行政院國家科學委員會補助專題研究計畫期末成果報告

支援下一代無線與 FTTx 擷取之光纖都會網路技術—子計畫四：
快速行動擷取網路中之品質服務支援技術研究
QOS Enabling Technology for High Mobility Wireless Network

計畫類別：□ 個別型計畫　　☑ 整合型計畫
計畫編號：NSC　92－2213－E－009－120－
執行期間：　92　年　8　月　1　日至　95　年　7　月　31　日

計畫主持人：陳耀宗
共同主持人：
計畫參與人員：　詹益禎、沈上翔、林政豪、施振華、劉頂立、林雋永

成果報告類型(依經費核定清單規定繳交)：□精簡報告　■完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：國立交通大學資訊工程學系

中　華　民　國　九十五　年　十　月　三十一日

「支援下一代無線與 FTTx 擷取之光纖都會網路技術」子計畫四
「快速行動擷取網路中之品質服務支援技術研究」
QOS Enabling Technology for High Mobility Wireless Network

計畫編號：NSC 92-2213-E-009-120

執行期限：92 年 8 月 1 日至 95 年 7 月 31 日

主持人：陳耀宗教授　(國立交通大學資訊工程學系)

計畫參與人員：詹益禎、沈上翔、林政豪、施振華、劉頂立、林雋永

一、中文摘要

本計畫在研究快速行動性(High mobility)下無線網際網路之品質服務(Quality of Services)技術，以能夠平順地支援無接縫式的即時多媒體串流(Real-time Multimedia Streaming)為目標。由於傳統IETF所制定之行動IP協定，是針對大範圍且跨不同網域之行動性而設，在低遲滯(Low Latency)之延展性(Scalability) 與效能(Efficiency) 方面不足以應付經常性且小範圍之行動管理需求，因此有不少針對微行動性(Micro mobility) 之協定被提出。這些協定之主要目的在減少訊號處理之負荷(Signaling Overhead)與降低由交遞(Handoff)造成之封包延遲(Packet delay)與封包遺失(Packet loss)。本計畫所欲解決之問題，除了上述各點之外，將針對快速移動之無線行動環境下之服務品質技術做深入之探討，包括了第三層路由器上訊號封包與資料封包之處理、路由更新(Routing update)訊號協定(Signaling Protocol)與封包轉送(Packet Forwarding)機制之設計，以及第二層交換器端在Handoff過程中針對品質服務提供(QoS Provisioning)之處理機制之研究。除此，本計畫也將與其它子計畫之實體層技術配合，以實際達到一個快速行動性下平順無接縫式之無線行動品質服務。我們從三方面進行研究：支援快速交遞(Fast Handoff) 之IP路由(Routing)機制設計，包括研究更有效率之訊號協定與支援快速行動性下QoS機制之路由方式研究設計；行動端位置管理(Location Management)，包括如何透過各種訊號方式有效地去偵測、追蹤與更新行動端之正確網路位址，並快速找出下一個擷取點，以利封包之傳送；以及快速交遞之支援，包括快速行動性下之交遞特性研究，與因應之訊號交換架構設計以達到最小化遲滯與封包遺失之快速交遞。在這三方面除了考慮品質服務之基本需求外我們也將延展性(Scalability)與系統效能列入研究重點。在計畫中，我們研究現行各種Micro mobility協定之優缺點並透過模擬系統做比較，而後提出研究改進後之方法。我們利用不同性質之訊務做品質服務模擬測試。針對UDP訊務，我們著重交遞過程中，封包延遲與封包遺失之最小化；而針對TCP訊務，我們儘量做到交遞過程中，流通率(Throughput)能夠不受影響。而後我們將所提之方法實現在網路設備與行動端上，以建立一品質服務無線網際網路平台，並導入實際之多媒體串流訊務做Real World 之測試。我們也研究不同無線擷取技術之間之交遞方式，並與其他子計畫做垂直技術整合。

支援無線區域網路行動性的無接縫快速交遞機制之部份研究中，均以交遞發生前先行偵測交遞對象為基礎，並已發展完整。鄰近圖(Neighbor graph) 是提供偵測交遞對象資訊的方法中最常見的一種。但鄰近圖因為先天性的一些缺點，至今還沒被廣泛的應用。在這計劃中，我們提出了所謂「不連續掃描」的機制來承襲鄰近圖的功能且排除其不實用的固有缺點。在不連續掃描的基礎上，我們更進一步發展了一些機制來幫助行動工作站在交遞前選擇合適的基地臺作為交遞對象，並用模型分析及模擬的方法來驗證這些機制的效能。「不連續掃描」最重要疑慮可能在於對行動工作站的服務品質產生衝擊。模擬結果顯示，對一個在無線區域網路 EDCA 上運作的媒體串流而言，不連續掃描造成的中斷將可以被控制在 50 毫秒以內。

我們也提出了一套兩階段認證機制以達到低遲滯交遞之目的。此機制容許行動裝置利用快速交遞協定送出一特定之認証資料以在漫遊至新的區域之前從新的擷取點獲得一臨時通行証。行動裝置可利用臨時通行証迅速地接收即時資料封包，並執行重新認証以完成完整的正常認証程序。我們評估所提方法之效能，它顯示了暫時認証機制改善了交遞程序的服務品質。

關鍵詞：快速行動性，無線網際網路，品質服務，即時多媒體串流，微行動性，快速交遞，鄰近圖，不連續掃描，暫時認證，臨時通行憑証，低遲滯交遞。

二、英文摘要

This project focuses on the investigation of QOS enabling technology under a high mobility wireless environment. It targets at smoothly supporting the seamless real-time multimedia streaming. Since the traditional Mobile IP protocol defined by IETF is focusing on a large scale and inter-domain mobility, its performance regarding the scalability of low latency and efficiency is insufficient to fulfill the requirement of managing the frequent and small scale mobility, called micro mobility, therefore many protocols focusing on micro mobility have been proposed. These protocols aim at reducing the signaling overhead and minimizing the packet delay, as well as packet loss caused by the fast handoff. In addition to these problems mentioned above, other issues to be solved in this project include the investigation of QOS-provisioning under high mobility environment, this involves the layer 3 packet processing for both signaling messages and data packets, signaling protocols for routing update, design of packet forwarding mechanism for fast handoff, and study of layer 2 mechanisms on switches regarding QOS related schemes under handoff. Further, we will cooperate with other subproject to work on the physical layer issues such as to achieve smooth and seamless QOS-enabled wireless mobile services. The project has been performed with three objectives: First, the design of IP routing mechanism to support fast handoff, this includes the investigation of a more efficient signaling protocol and QOS-capable mechanism inside routing equipments; second, location management of mobile hosts, this includes how to use various signaling messages to effectively detect, track and update the IP address of a mobile host precisely, so that packets can be forwarded efficiently and correctly during fast handoff period; and third, handling of the fast handoff, this includes the investigation of handoff characteristics under high mobility, and the corresponding signaling exchange infrastructure so that latency and packet loss can be minimized during the handoff. We also considered both scalability and performance issues in addition to the three objectives described above. We started from the study of existed micro mobility protocols, then we compared their pros and cons through system simulation. We performed the simulation regarding the QOS evaluation using data traffic with different characteristics. For UDP traffic, we will focus on the minimization of packet delay and packet loss during fast handoff; while for TCP traffic, we emphasized on the stable throughput during the handoff under high mobility. We realized our proposed approach on the network equipment and mobile devices, so that we could build a QOS-enabled wireless Internet platform, on which the real-time multimedia streaming traffic could be added on for the experiment and system evaluation. We also investigated the fast handoff protocols between various wireless access technologies, and make vertical technology integration with other subjects.

Seamless or fast handoff schemes to support mobility of IEEE 802.11 Wireless LANs have been well developed based on the supposition that a set of next-AP/AR candidates have been available prior to a handoff. A neighbor graph is one of the major strategies used to provide such information among the related studies. However, neighbor graphs do not been widely applied yet due to its inherent drawbacks. In this project, a so-called “discrete scan” scheme is designed to substitute for the function of a neighbor graph without its drawbacks. Moreover, several mechanisms based on discrete scan are presented to help a mobile node select a desired next AP to handoff to. The analytical model and simulation are elaborated to show performance of the approaches. Discrete scan may bring concerns on its impact to received QoS of a mobile node. However, the simulation results show that disruptions caused by discrete scan are controlled less than 50 ms for a media streaming device working on EDCA over wireless LAN.

We also propose a two-stage authentication scheme in this project for achieving low-latency handoff. It allows mobile devices to send certain authentication information by fast handover protocol to obtain a temporary pass certificate from new access node before roaming to the new domain. Mobile device can use the temporary pass certificate to receive real-time data packet quickly and then perform re authentication process to complete the total authentication as normal procedure. We evaluate the performance of the proposed method and it is shown that transient authentication scheme really improve the quality of service during handover process.

Keywords: High mobility, wireless Internet, Quality of Services, Real-time multimedia streaming, micro mobility, fast handoff, neighbor graph, discrete scan, transient authentication, temporary certificate, low latency handoff.

三、計畫緣由與目的

The ever-growing demand for Internet bandwidth and recent advances in optical Wavelength Division Multiplexing (WDM) and wireless technologies brings about fundamental changes in the design and implementation of the next generation networks. To support end-to-end data transport, there are three types of networks: wide-area long-haul backbone network, metropolitan core network, and local and access networks. First, due to steady traffic resulting from high degree of multiplexing, next-generation long-haul networks are based on the Optical Circuit Switching (OCS) technology by simply making relatively static WDM channel utilization. Second, a metropolitan core network behaves as transitional bandwidth distributors between the optical Internet and access networks. Unlike long-haul backbone networks, metro networks exhibit highly dynamic traffic demand, rendering static WDM channel utilization completely infeasible. Finally, access networks are responsible for providing bandwidth directly to end-users. Two most promising technologies have been optical access and wireless access networks, respectively. Due to superior performance of fiber optics and tremendous bandwidth demand, providing broadband access and services through optical access technology becomes indispensable. Finally, regarding wireless access networks, the new demand of wireless communications in recent years inspires a quick advance in wireless transmission technology. Technology blossoms in both high-mobility low-bit- rate and low-mobility high-bit-rate transmissions. Apparently, the next challenge in wireless communications would be to reach high transmission rate under high mobility.

The main objective of this subproject is the provision of QoS guarantees over wireless access networks. By investigating the QOS enabling technology under a high mobility wireless environment, we attempt to smoothly support the seamless real-time multimedia streaming.

四、研究方法與成果

The subproject is performed with three directions: first, the design of IP routing mechanism to support fast handoff; second, location management of mobile hosts; and third, handling of the fast handover through two-stage authentication; In the first part, we developed a new handover scheme named “speedy handover” to enhance the performance of wireless handover. The proposed scheme makes use of IEEE 802.11 [1] RTS/CTS exchanging messages to quickly detect the movement of mobile nodes. It can improve the performance of traffic transmission during the handover period. In the second part, we investigated a so-called “discrete scan” scheme to substitute for the function of a neighbor graph. Several mechanisms based on discrete scan are proposed to help a mobile node select a desired next Access Point to handoff to. The simulation results show that disruptions caused by discrete scan are controlled less than 50 ms for a media streaming device working on EDCA over wireless LAN. In the third part, we investigated a low-latency handover scheme through a so-called two-stage authentication, because authentication process contributes most of the inter-domain handover latency. These works are discussed as follows:

1. Speedy Handover

1.1 Introduction

During handover period, packets for the mobile node may be lost in its old foreign agent (FA), because it will attach to another new FA and has detached from the old one. We want to keep these packets and forward them to the destined mobile node. However, when to buffer and to forward packets are critical issues which require further investigation.

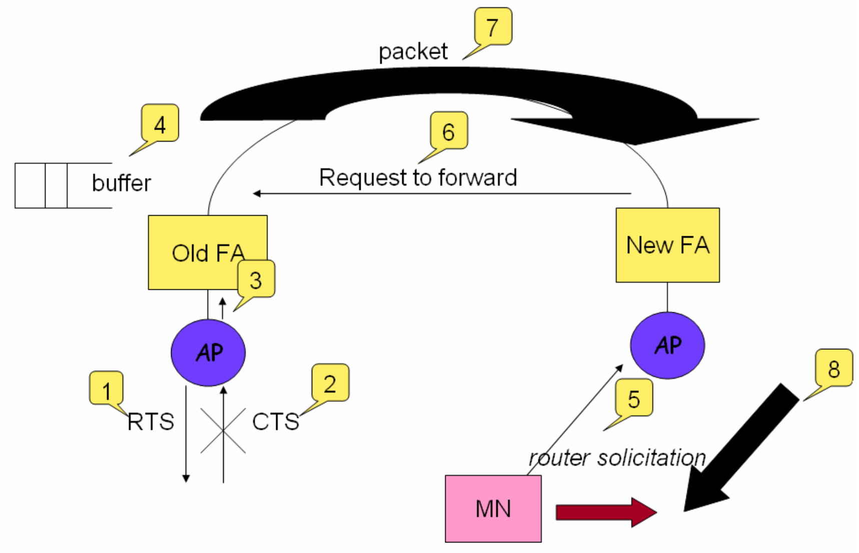


Fig. 1: Messages flow of speedy handover scheme.

In our scheme, we use RTS/CTS messages exchange between FA and the mobile node to detect whether a mobile node still attaches to the FA or not. The RTS/CTS messages exchange is an important method for solving hidden terminal problem in IEEE 802.11. When a FA wants to send a data packet to a mobile node, it sends a RTS message. Upon receiving a CTS message from the mobile node, the FA starts to send data packets. Since RTS/CTS messages are short, frequently transmitted, and less affected by random loss, thus we can use them to detect the mobile node movement.

1.2 Description of the Scheme

The proposed speedy handover mechanism can be briefly described by Fig. 1. When a FA has sent RTS for three times but not being responded by the mobile node, we can infer that the mobile node has moved away. It is a good time point to start buffering the packets for the mobile node. Steps 1 and 2 show that the AP (access point) of the old FA sends RTS three times and no CTS responded by the mobile node. So the AP sends a specific message to the old FA to request the FA to buffer packets for the mobile node, as depicted in steps 3 and 4.

When a mobile node gets the beacon from the new AP, it delivers “router solicitation” messages. The message informs the new FA that a new mobile node has arrived, as shown in step 5. Then the new FA sends “request to forward” message to old FA as soon as the routing table to mobile node is updated, as shown in step 6. Finally, in steps 7 and 8, the old FA forwards packets to the mobile node through the new FA.

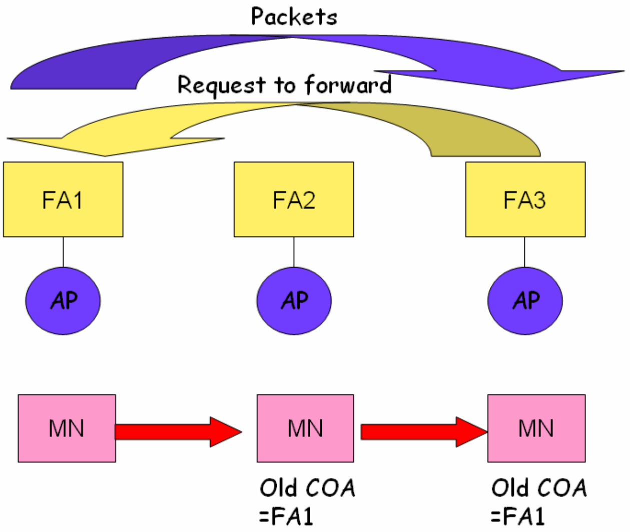


Fig. 2: Fast movement of a mobile node (scenario 1).

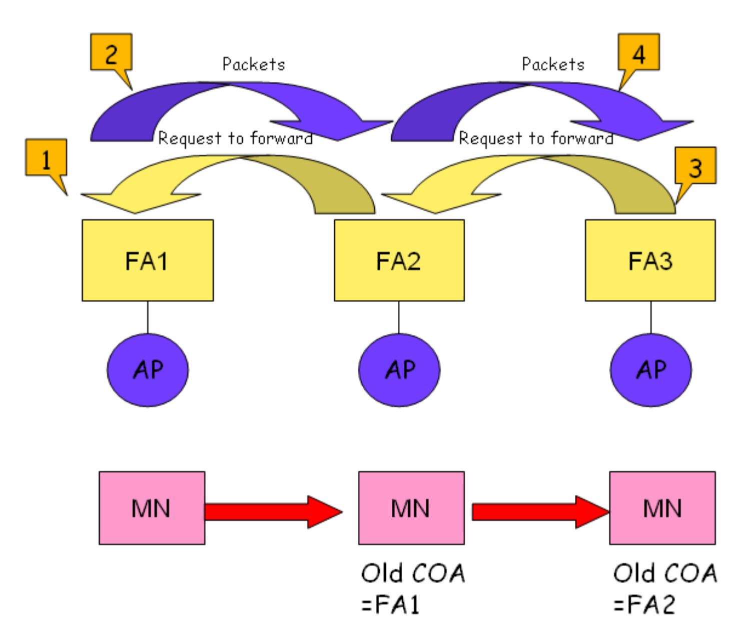


Fig. 3: Fast movement of a mobile node (scenario 2).

Speedy handover needs a simple modification to a mobile node. Since the mobile node wants the new FA to send the “request to forward” message to the old FA, thus the new FA must know where the mobile node comes from. Therefore the old care of address (COA) of the mobile node will be added in router solicitation messages.

Speedy handover can still work when the mobile node moves fast. As shown in Fig. 2, the mobile node leaves the FA2 very quickly before router solicitation is send. After the mobile node reaches the FA3, the old COA (care of address) record in it is the IP of the FA1. Therefore, the FA3 will send “request to forward” message to the FA1. Since the mobile node does not complete the handover to the FA2, so packets for the mobile node are still sent to the FA1 before handover to the FA3 finish. FA1 keeps packets for the mobile node and then forwards them to the FA3.

Another fast movement scenario is depicted in Fig. 3, the mobile node leaves the FA2 after router solicitation is sent but before the handover finishes. Because of the router solicitation, packets for the mobile node are forwarded to the FA2 from the FA1. After the mobile node reaches the FA3, the FA3 send “request to forward” message to the FA2. Therefore packets for the mobile node are forwarded from the FA1 to the FA2 and then from the FA2 to the FA3. The mobile node is able to receive the packets successfully.

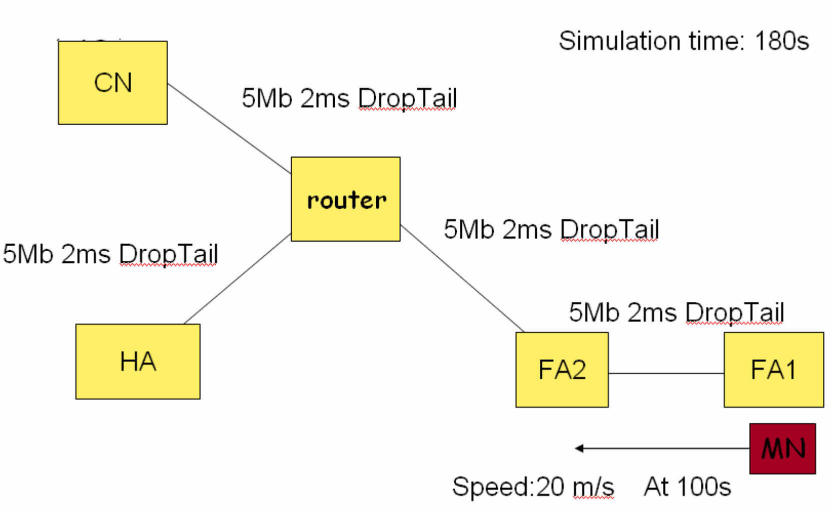


Fig. 4: Network topology for simulations.

1.3 Performance evaluation

We evaluate the proposed scheme with network simulator ns2 [2] and the simulation network topology is shown in Fig. 4. The performance of mobile node during handover period is what we want to know. There are one mobile node (MN), one corresponding node (CN), two foreign agents (FA), and one home agent (HA) in the simulation network. The communication range of all nodes with wireless interface is 550 m. The distance between FA1 and FA2 is 856 m and the overlap of communication range is about 244 m. The simulation starts at 0 second and ends at 180 second. An UDP sender (CN) starts to send packets at 100 second until the end of simulation and the mobile node begins to move from the FA1 to the FA2 at the same time with speed 20 m/s. The packet size is 1000 bytes and sending rate is 1 Mb/s. We give each packet a sequence number. By checking the packet sequence numbers we can observe that which packet is received by the mobile node and which packet is lost. We compare the performance between the original structure and speedy handover.

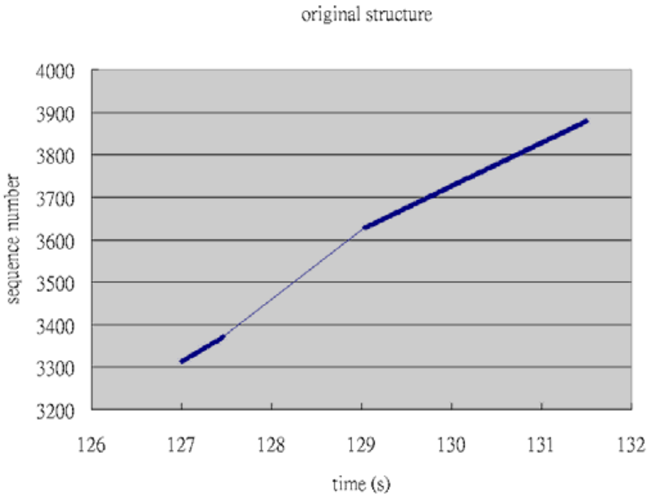


Fig. 5: Sequence number of packets that the mobile node receives with original structure.

The handover begins at about 127.46 second, so we show the result from 127 second to 131.5 second. As shown in Fig. 5, the mobile node can not receive any packets from the UDP sender (CN) during the handover period. Packets are lost because no body buffers and forwards them for the mobile node. Obviously the packet loss rate is very high.

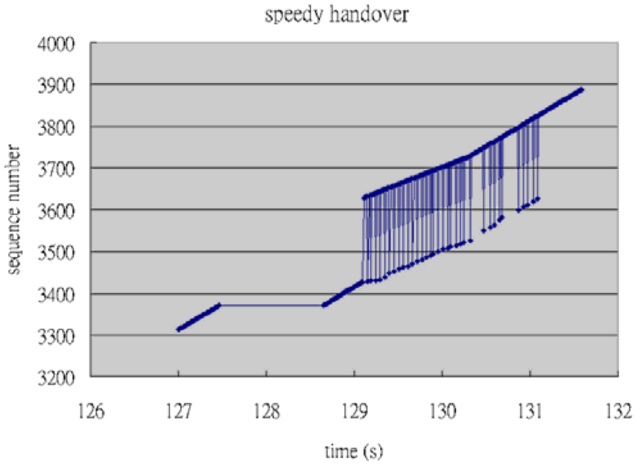


Fig. 6: Sequence numbers of packets that the mobile node receives with speedy handover.

Figure 6 depicts the performance when the speedy handover mechanism is used. The old FA buffers the packets during handover period and forwards them as soon as the new FA knows how to deliver the packets to the mobile node. The mobile node can receive the packets before handover procedure finishes. The handover procedure is complete at about 129.15 second and the mobile receives forwarded packets at about 128.66 second, so we shorten the time that the mobile node can not receive any packets. After handover procedure finish the new packets stream toward the mobile node make the packets out of order. It is the cause of the wavy line in Fig. 6.

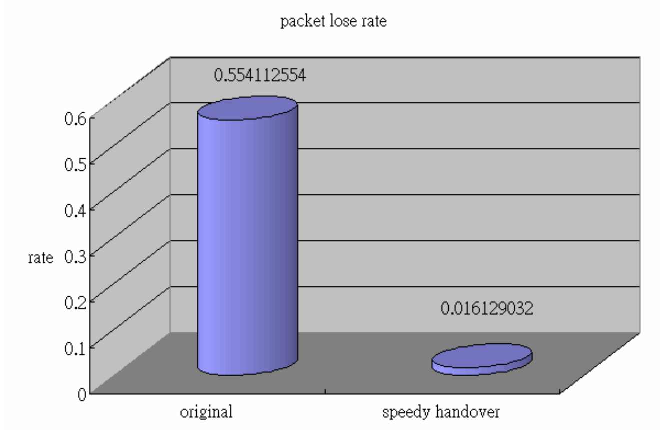


Fig. 7: Packet lose rate between handover start to forward finish.

By observing the Fig. 7, we can find that the packet loss rate for speedy handover is much lower than that of original structure. The packet loss rates are 55.41% and 1.61% for the original structure and speedy handover respectively. For a multimedia service, low packet loss rate is an important condition for smooth quality.

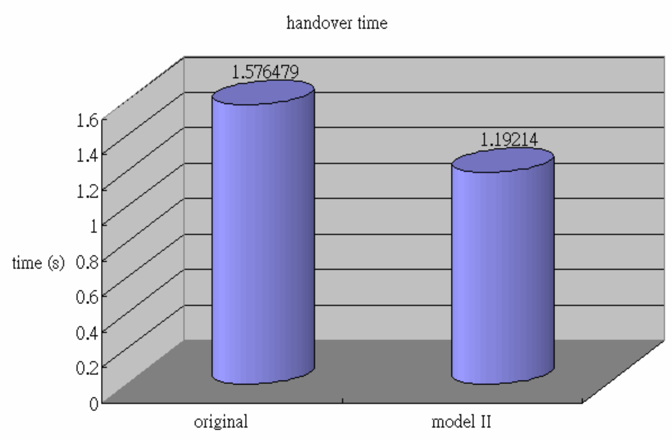


Fig. 8: Handover time.

Figure 8 shows the time from the start of handover to end of period that the mobile node can not receive any packets. It is clear that speedy handover can shorten the time.

1.4 Section Conclusion

In short, comparing with the original mobile IP mechanism, speedy handover features a fewer number of packet losses and a shorter handover period. The simulation results demonstrate the effectiveness of the proposed scheme. In this chapter, we propose a so-called “discrete scan” scheme and relative auxiliary mechanisms that collect a set of candidates for next-APs and eventually select the next AP for a mobile node before a forthcoming handoff. Moreover, various existed mechanisms may select next AP with different properties of the next-AP candidates, such as the nearest AP, the approached AP, the AP of highest available bandwidth, or their combination.

2. Discrete Scan

2.1 Introduction

In discrete scan, a mobile node, such as a WiFi phone set, will make use of the idle time in pre-handoff duration to sniff its wireless environment to collect a set of next-AP candidates for a forthcoming handoff. A pre-handoff period is initiated with an event that received signal strength (RSS) from current AP is detected to be lower than a preset threshold for a preset period.

Figure 9 below illustrates the typical scheme for the wireless NIC in the pre-handoff periods. While working in pre-handoff mode, the wireless NIC has to return to working channel to maintain VoIP connection alive with a certain level of QoS. The time between two working periods, that is, the time of sniffing period plus two switching periods, should not be longer than β ms so that the maximum latency allowed by a real time application, α ms, is maintained if the node will get at least two times of transmission within (α−β) ms, as shown in Fig. 9. During working periods, bidirectional traffic transmits between current AP and the mobile node at least once to deliver the traffics generated in previous absence from working channel. By the end of working periods, the mobile node should issue an extra control frame of power save mode (PSM) to notice the current AP to suspend the packets for the mobile node.

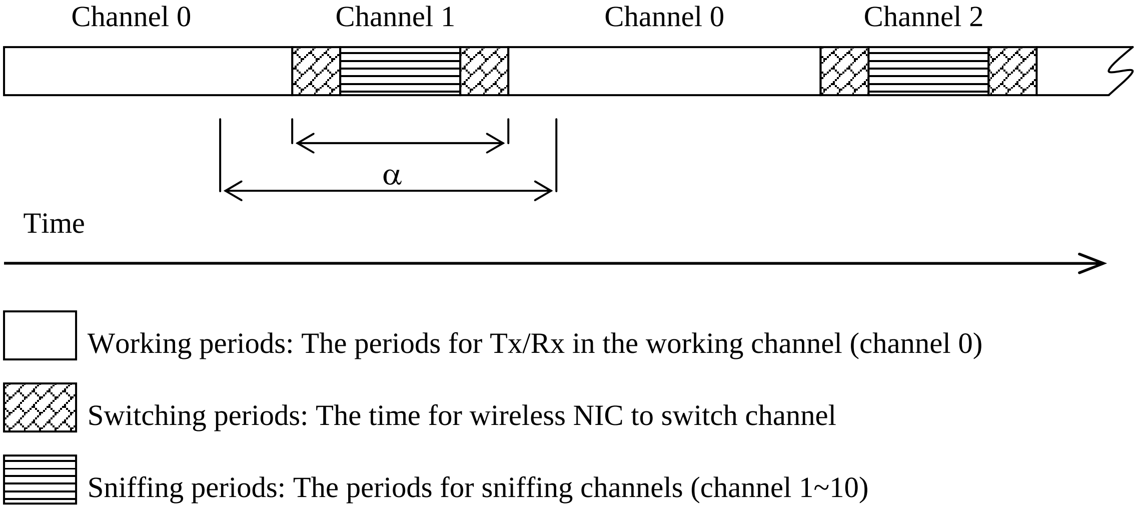

Channel 0
Channel 1
Channel 0
Channel 2
α
Time
Working periods: The periods for Tx/Rx in the working channel (channel 0)
Switching periods: The time for wireless NIC to switch channel
Sniffing periods: The periods for sniffing channels (channel 1~10)

Fig. 9: Scheme for wireless NIC in pre-handoff periods.

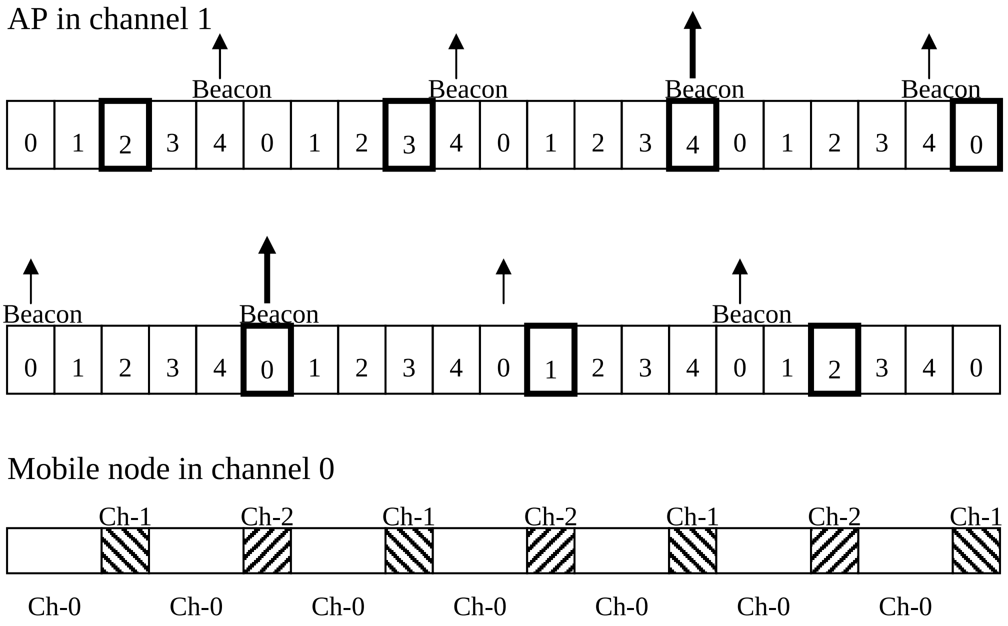


Fig. 10: Example sniffing scheme in the environment of 3 channels.

To substitute for the function of neighbor graphs, discrete scan must be able to collect a set of next-AP candidates. One of the nature particular to a seamless handoff is that a mobile node must have entered the coverage area of its next AP before a layer-2 handoff is triggered. Therefore, a mobile node can discover the entire set of next-AP candidates by means of sniffing frames transmitted from APs in various channels. An example design for the sniffing scheme is illustrated in Fig. 10. A sniffing period is designated by 20 ms for each interval of 60 ms. Assume that only three channels are deployed in the concerned area. A beacon interval consists of 5 slots, from slot 0 to slot 4, with slot time 20 ms, and two channels which are different from the working channel will be listened in a sequence of sniffed periods. The sniffed channels should be interleaved so that each channel is probed evenly.

The maximal time required for a mobile node to discover all APs through beacons should be taken into account to set the value of δ_2, the threshold RSS to start a pre-handoff period. As an example, 20 ms slot of each 60 ms interval is used for sniffing, that is, 1/3 of total run time is shared to sniff for the next-AP candidates in all channels except the working channel. Each scanned channel takes 5 sniffing periods, equal to 100 ms in total, to ensure to discover of all of the next-AP candidates active in the channel. Therefore, it takes (number of channels–1) * 3 * 100 ms to complete the discrete scan scheme on all of channels. With at the most 11 channels in IEEE 802.11b wireless LAN environment, thus, it takes (11-1) * 3 * 100 ms, equal to 3 sec, to complete a full passive scan. Therefore, a pre-handoff period is suggested to be 3 seconds ahead of a layer-2 handoff.

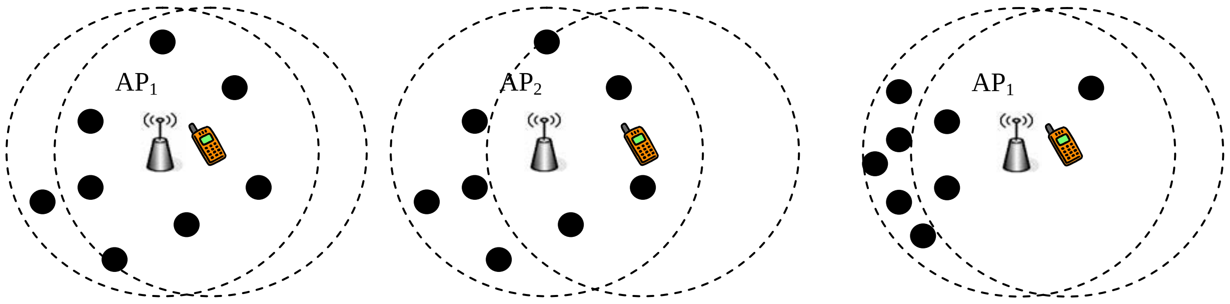
2.2 Mechanisms to select the nearest AP

Discrete scan surpasses the neighbor graph in the ability to select a proper next AP for a mobile node before the handoff starts. Taking the advantage of the information extracted from MAC frame headers that a mobile node listens through its discrete scan scheme; a desired next AP can be determined by the mobile node along with extra supports from existed wireless LAN system. The feature to determine the next AP may take over the function of a neighbor graph with aids of GPS as proposed in [4].

In a conventional handoff schemes, a mobile node determines the nearest AP by the RSS received from beacons or frames sent by APs during the probe phase of a layer-2 handoff. However, it is recognized that the strongest RSS may be not a prefect indicator of the nearest AP because of the effects by multi-paths interference of microwave communications. Nevertheless, the principle of choosing the next AP by indication of RSS is widely implemented in wireless NIC, since no better indicator is available so far. Discrete scan, discovering next-AP candidates through frames or beacons sent by APs, may restrict the capability to select next AP by their RSS. Furthermore, several mechanisms of a new idea with aids of MAC layer information are proposed in this section to compensate the insufficiency of RSS scheme currently supported by physical layer.

By reading MAC layer headers, a mechanism is designed to record the number of stations that can be listened by the mobile node in each channel. A mobile node learns a station with its MAC address in the frame header transmitted from the station.

The first proposed mechanism for a mobile node to estimate its distances from candidate APs is named as “station distribution ratio” that is defined as the ratio of the numbers of stations of in the coverage of the mobile node to all of the stations in a BSS. As illustrated in Fig. 11(a) (b), the higher station distribution ratio is a BSS, the nearer is the AP to the mobile node.



High station distribution ratio (7/8) indicates a near AP.

(a)

Low station distribution ratio (4/8) indicates a far AP.

(b)

The cluster of station leads low station distribution ratio (3/8) for a near AP

(c)

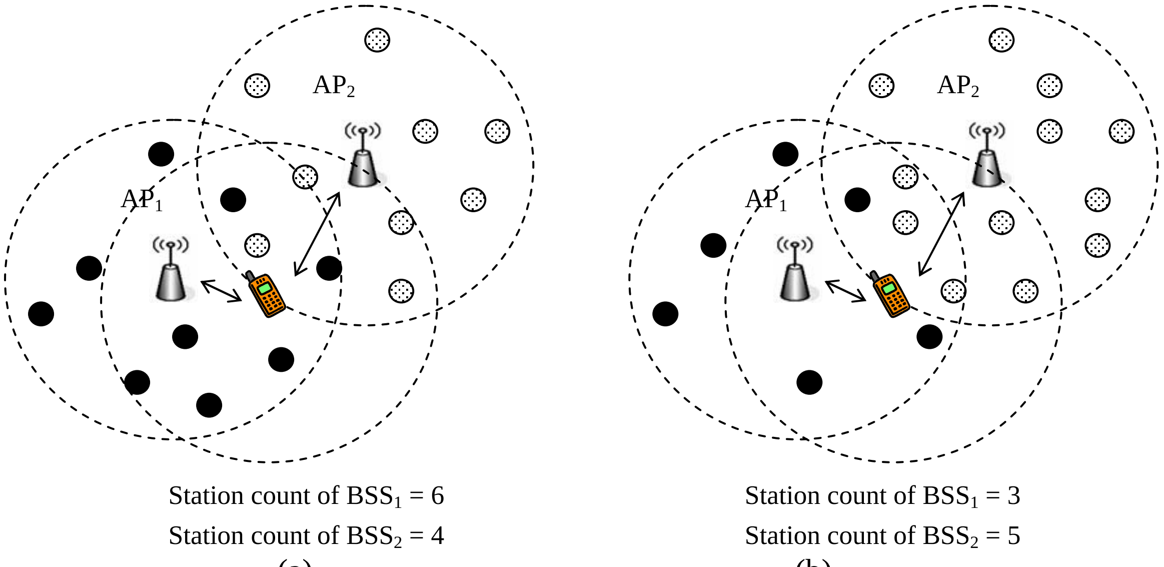
Fig. 11: the higher station distribution ratio indicates the nearer AP.

A mobile node can identify a station in its coverage area if the mobile node can receive both a data frame and the corresponding ACK frame in its discrete scan scheme. Similarly, if a mobile node receives only one of a data frame or the corresponding ACK frame, the mobile node infers that the station in the BSS locates out of its coverage. Another good feature that discrete scan surpasses neighbor graphs is its slight penalty if an incorrect AP is selected by the auxiliary mechanisms. A mobile node always locates at the overlapped coverage of all of the next-AP candidates discovered in the discrete scan scheme, therefore, an improper selection among the next-AP candidates will not lead to a failure of a handoff. On the contrary, a mobile node will lose connection if the GPS in a mobile node predicts a neighboring AP that can not reach to the mobile node as the next AP. As illustrated in Fig. 11(c), the cluster of stations in a BSS may lead the station distribution ratio mechanism to an improper selection of the next AP. However, the penalty for a mobile node to handoff to a farer AP may just cause another handoff to occur earlier.

There are some shortcomings of the station distribution ratio mechanism. It will be active only after an AP is discovered because a mobile node infers a station out of its coverage by frames from AP instead of the station. Consequently, the station distribution ratio mechanism requires a certain amount of time to evaluate an AP after it is discovered. Besides, a mobile node needs more extra cache to memorize the MAC addresses of discovered stations to avoid counting a station twice.

To overcome the above drawbacks, the station distribution ratio mechanism is simplified and submitted as the second mechanism for a mobile node to select the nearest AP among candidate APs. The mechanism is named as “station count” because it suggests a mobile node to select an AP of the most number of stations counted by a mobile node in the last sniffing period prior to a handoff. A mobile node with station count mechanism updates the number of stations it found in a BSS in each short sniffing period (e.g., 20 ms) of a discrete scan scheme. Because the mobile node concerns only the stations that locate in its coverage area and transmit at least one frame in a short sniffing period, much less cache memory is needed to keep MAC addresses of discovered stations, and a simple algorithm that collects the frames of various senders to a receiver is sufficient to identify an active station in the coverage. Furthermore, a mobile node starts to sense a BSS before it enters the coverage; therefore the evaluation of a BSS is available with discrete scan once its AP is discovered by the mobile node.

As illustrated in Fig. 12(a), the nearer is the mobile node to an AP, the larger overlapped coverage the mobile node can listen, and consequently the mobile node can discover more active stations in a sniffing period. Unfortunately, the distance between a mobile node and an AP is not a unique factor that affects the station counts. As illustrated in Fig. 12(b), the total number of stations in a BSS is another major factor that affects the station counts in a sniffing period. Due to the imbalance of number of stations in two BSSs, the far AP may result in larger station count than a near one.



Station count of BSS$_1$ = 6
Station count of BSS$_2$ = 4
(a)

Station count of BSS$_1$ = 3
Station count of BSS$_2$ = 5
(b)

Fig. 12: The station counts in two BSSs.

The third factor that affects the station counts in a sniffing period is the length of a sniffing period. The longer sniffing time results in larger station count until the mobile node discovers all of the stations in its coverage. On the other hand, if the sniffing period is as short as a slot time in DCF wireless environment, the station count is reduced to the probability for one of the stations in the mobile node's coverage to get a successful transmission in the given timeslot. Since all of the stations in a BSS share a media of limited bandwidth, a BSS of fewer stations has higher probability to transmit a frame in an arbitrary timeslot because of less collisions and smaller contention windows. Hence, a short sniffing period benefits a BSS containing fewer stations.

The length of sniffing periods of discrete scan is properly set to either balance or alleviate the affect of the station counts brought by imbalance of number of stations in various BSSs. There are two major concerns on setting the length of a sniffing period: first, a long sniffing period may eventually eliminate the affect brought by imbalanced number of stations. Second, a short sniffing time results in insufficient station counts to have creditable selection among the candidate APs.

In the following, we will elaborate the way to determine an appropriate length of a sniffing period. The model proposed by G. Bianchi [5] is employed for the analysis of DCF wireless LAN environment.

2.3 Setting of a sniffing period

As described before and illustrated in Fig. 12(a) (b), two factors control the station count of a BSS, distances between the mobile node and the AP and number of stations in the BSS. Intuitively, we have

$$E(s) \approx r \cdot A' / A \qquad (1)$$

where *E(s)* denotes the expected value for *s*, and s denotes the station count of a BSS in a sniffing period that is represented by T_{sniff}; *r* is a new term called as "number of transient stations", that is defined as the number of stations which at least send one frame in given sniffing time (T_{sniff}) in a BSS; and *A'* denotes the overlapped coverage of a BSS and the mobile node, and *A* denotes the coverage area of BSS.

Let τ denote the probability that a station transmits in an arbitrary slot time, and *p* denotes the probability that a transmitted packet encounters a collision. By [5], in a BSS of DCF wireless environment, there are

$$\tau = \frac{2(1-2p)}{(1-2p)(W+1)+pW(1-(2p)^m)} \qquad (2)$$

$$p = 1-(1-\tau)^{n-1} \qquad (3)$$

where, *n* is the number of stations in a BSS
W is the minimal contention window in DCF.
m is the integer such that $2^m W$ presents the maximal contention window in DCF.

For a given *n*, τ and *p* are derived from (2) and (3). Thus, the average number of frames transmitted within one timeslot would be τ∗n.

Let P_{tr} denote the probability that at least one frame is sent in a considered timeslot. There are *n* stations and each station transmits a frame with probability τ, therefore,

$$P_{tr} = 1-(1-\tau)^n \qquad (4)$$

Let P_s denote the probability of a successful transmission in a considered timeslot, i.e., the probability that exactly one station transmits on the channel, under the condition of the fact that at least one station transmits. It has,

$$P_s = \frac{n\tau(1-\tau)^{n-1}}{P_{tr}} = \frac{n\tau(1-\tau)^{n-1}}{1-(1-\tau)^n} \qquad (5)$$

Let σ be a slot time in the condition of idle media. T_s denotes the length of a slot time in the condition of a successful transmission and T_c denotes the length of a slot time in the condition of a transmission with collision, as defined in Fig. 13.

$$T_s = PHYhdr + MAChdr + E[P] + SIFS + \delta + ACK + DIFS + \delta \qquad (6)$$

$$T_c = PHYhdr + MAChdr + E[P^*] + DIFS + \delta \qquad (7)$$

where, δ stands for the propagation delay, *E[P]* is the average length of packet payload and *E[P*]* is the length of the longest packet payload involved in a collision.

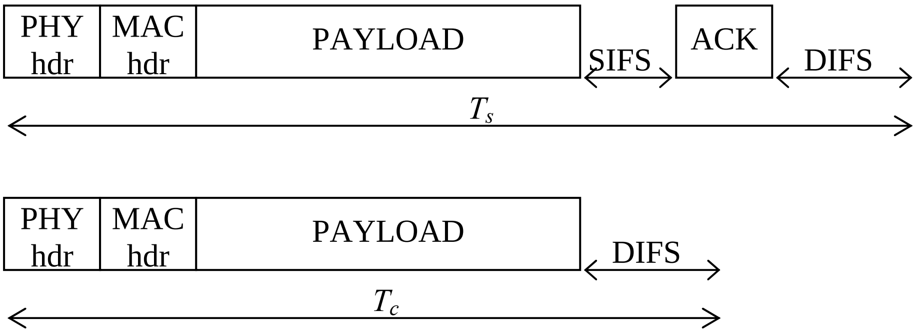


Fig. 13: Definition of T_s and T_c, slot time of transmission with success and collision.

The average length of a slot time, denoted by T_{av}, is readily obtained by considering that, with probability (1-P_{tr}), the slot is null of transmission; with probability P_{tr} P_s, the slot contains a successful transmission, and with probability P_{tr}(1-P_{tr}), the slot contains a collision. Hence,

$$T_{av} = (1 - P_{tr})\sigma + P_{tr}P_s T_s + P_{tr}(1 - P_s)T_c \qquad (8)$$

Let *m* denote the total number of frames sent by *n* stations within T_{sniff}. It is estimated in average as,

$$m = n\tau * T_{sniff} / T_{av} \qquad (9)$$

Parameters	Values
PLCP preamble & header	192.0 μs
MAC header	20.4 μs
Slot time	20 μs
SIFS	10 μs
DIFS	50 μs
ACK	10.2 μs + PLCP preamble & header
CW_{min}	32
CW_{max}	1023
Maximal Packet payload	2312 bytes
Channel bit rate	11 Mbps

Table 1 Parameters for IEEE 802.11b wireless LAN.

Assume that *n* stations transmit *m* frames randomly. The number of transient stations, *r*, is the number of stations that send at the least one of *m* frames in T_{sniff}. Thus, *r* is derived as follows:

$$r = \frac{\sum_{k=1}^{n} k \cdot P(n,k) \cdot S(m,k)}{\sum_{k=1}^{n} P(n,k) \cdot S(m,k)} \tag{10}$$

where, *S(m,n)* denotes the Stirling number of the second kind, as,

$$S(m,n) = \frac{1}{n!} \sum_{k=0}^{n} (-1)^k \binom{n}{n-k} (n-k)^m \tag{11}$$

and *P(n,k)* denotes the permutation number as,

$$P(n,k) = \frac{n!}{(n-k)!} \tag{12}$$

Note that *m* in (9) is not an integer but *m* in (10) and (11) is restricted to an integer. Interpolation will be used in the numerical approach below.

With equations (2)~(12) and parameters listed in Table 1 for an IEEE 802.11b wireless LAN, the relations of transient number of station*(r)* v.s. number of station*(n)* of various T_{sniff} values are demonstrated in Fig. 14.

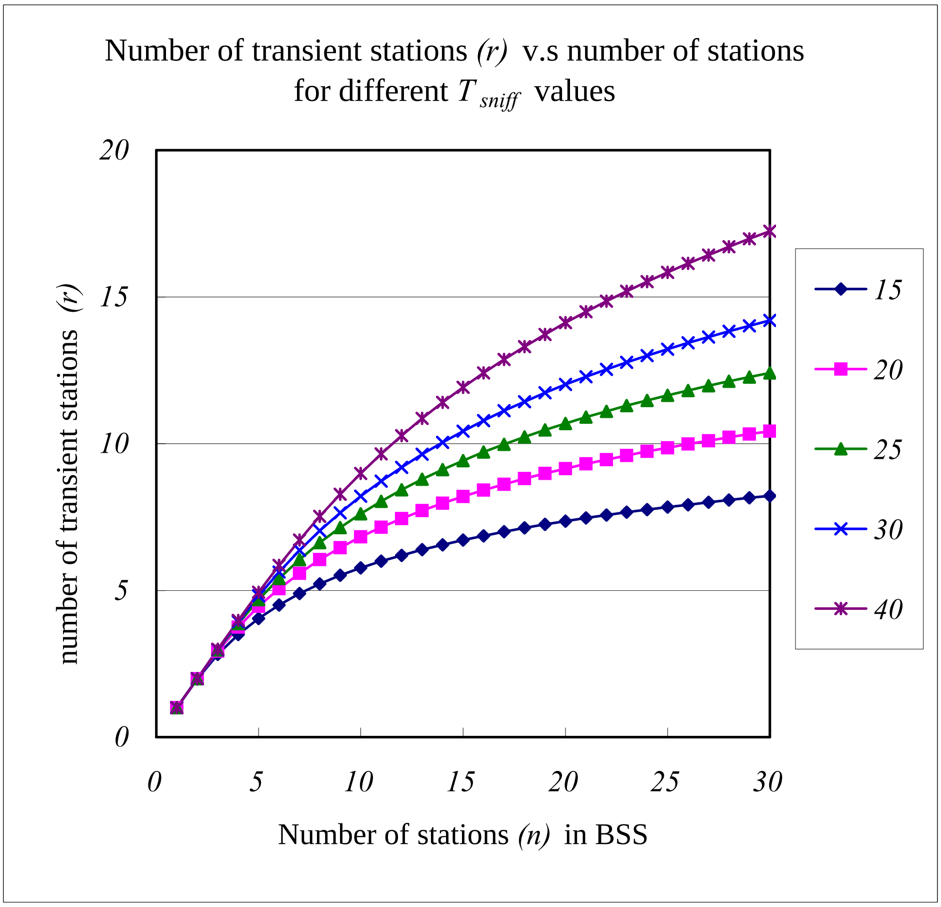


Fig. 14: Number of transient stations *(r)* vs number of stations *(n)* for different T_{sniff} values.

By reviewing equation (1), a station count (*s*) linearly depends on overlapped area (*A'*) and transient number of station(*r*) with respect to a T_{sniff}. Since it is desired that a station count to indicates overlapped area *(A')* rather than being disturbed by the imbalance of numbers of stations in different BSSs, an appropriate sniffing period *(T_{sniff})* should be set so that the number of transient stations *(r)* behaves insensitively to variation of the number of stations *(n)*.

Figure 14 shows the fact that a shorter sniffing period *(T_{sniff})* responds to a number of transient stations *(r)* less insensitive to variation of number of stations *(n)*. However, as described in the previous section, a sniffing period should be long enough to make station counts big enough for a creditable selection. In this project, we recommend that in a BSS of fewer stations (such as n = 6), number of transient stations *(r)* should be 80 % of number of stations *(n)*, i.e., r > 4.8 as n = 6. By the curves in Fig. 14, a sniffing period T_{sniff} is eventually chosen as 20 ms because the curve of T_{sniff} = 20 ms shows that its transient stations equal to 5 at n = 6. As n = 24, the number of stations increases by 4 times, the number of transient stations (r) is approximately equal to 10, increased by only 2 times. The numerical result shows that the variation of number of stations in a BSS reflects only half of numbers of transient station by setting a sniffing period as 20 ms. Although there is no proper sniffing period to make number of transient stations independent of number of station, however, the effect brought by imbalance of number of stations has been significantly mitigated by 50%

2. 4 Performance of Station Count Mechanism

E(s), the expected value of the station count with respect to a given T_{sniff}, is approximately equal to *r*A'/A* as described in (1), while collisions within overlapped area, *A'*, is neglected. In this section, the collisions are taken into consideration for a more accurate formula to estimate *E(s)*.

The average number of stations lies in the overlapped coverage that is estimated by area proportion, so that *n' = n*A'/A.* Let *m'* denote the total frames sent by either one of *n'* stations (no collision with any frame sent by one of *n'* stations) in T_{sniff}, then *m'* is estimated in (13), where τ is given in (2) and *m* is given in (9)

$$m' = n'\tau(1-\tau)^{n'-1}\cdot T_{sniff}/T_{av} = m\cdot(1-\tau)^{n'-1}\cdot n'/n \qquad (13)$$

Same as the process to estimate the number of transient stations, assuming that *n'* stations transmit *m'* frames randomly and let μ denote the reformed station count, the number of stations that sends at least one of *m'* frames in T_{sniff}. Thus, μ can be derived using (14) below.

$$\mu = \frac{\sum_{k=1}^{n} k\cdot P(n',k)\cdot S(m',k)}{\sum_{k=1}^{n} P(n',k)\cdot S(m',k)} \qquad (14)$$

where, *S(m,n)* denotes the Stirling number of the second kind that is defined in (11), and *P(n,k)* denotes permutation number that is defined in (12)

Note that *m'* in (13) may not be an integer and *m'* in (14) is restricted to an integer. Interpolation is used for numerical approach below.

Let S_i denote the random variable of station count in a sniffing period T_{sniff} for i^{th} channel. It is reasonably assumed that S_i features Poisson distribution and μ_i derived in (14), an estimation of reformed station count for i^{th} channel, should be the average of random variable S_i, i.e., $\mu_i = E(s_i)$.

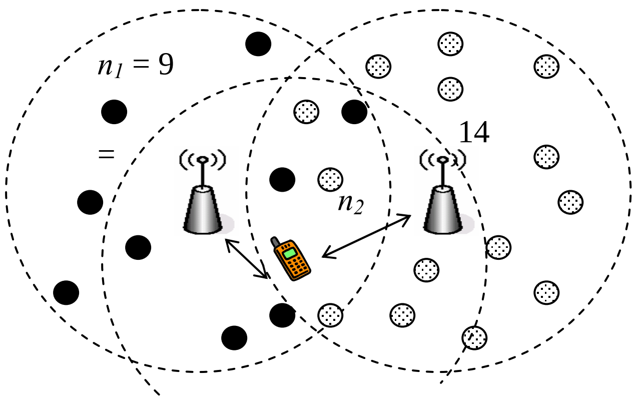


Fig. 15: Scenario for performance estimation.

The scenario to display the performance of station count mechanism is illustrated in Fig. 15. It assumes that a mobile node with discrete scan scheme roams in the overlapped coverage of two next-AP candidates. The current AP is not shown in Fig. 15, for its location will be irrelevant to performance of the station count mechanism. The mobile node selects one of two next-AP candidates, AP$_1$ and AP$_2$, according to their station counts derived in the last two sniffing periods of the discrete scan. Let *h* denote the hit ratio, the probability that the mobile node selects the nearest AP with the indication of larger station count as its next AP. Therefore, the mathematical definition for *h* in the case of two candidates is given in (15).

$$h = p(s_1 > s_2 \mid d_1 < d_2) + \frac{1}{2} p(s_1 = s_2 \mid d_1 < d_2) \qquad (15)$$

where $p(\ldots|\ldots)$ denotes the conditioned probability; s_i denotes the station count in T_{sniff} of i^{th} channel; and d_i denotes the distance between the mobile node and the AP in channel *i* (denoted by AP$_i$).

The average station count in each channel, μ_1 and μ_2, is estimated by (14) with given n_1, n_2, d_1, d_2, T_{sniff} as wells as parameters for DCF wireless environment (listed in Table 1). Since S$_1$ and S$_2$ are in Poisson distribution of mean values $E(s_1) = \mu_1$ and $E(s_2) = \mu_2$, the hit ratio *h*, defined in (15) is estimated by,

$$\begin{aligned} h &= \sum_{k=1}^{\infty} p(s_1 = k) \cdot [p(s_2 < k) + \frac{1}{2} p(s_2 = k)] \\ &= e^{-(\mu_1+\mu_2)} \sum_{k=1}^{\infty} [\frac{\mu_1^k}{k!} (\sum_{l=0}^{k-1} \frac{\mu_2^l}{l!} + \frac{\mu_2^k}{2k!})] \end{aligned} \qquad (16)$$

Hit ratios are demonstrated for the conditions that the mobile node locates at the place $d_1 = 0.5R$ and d_2 varies from $0.5R$ to $1.2R$, where *R* denotes the radius of coverage of BSSs and the mobile node. T_{sniff} is set as 20 ms. In the first case, it demonstrates how the distances between the mobile node and APs affect the hit ratio by setting the same number of stations in BSSs to ignore effects of the imbalance of numbers of stations in BSSs. In Fig. 16, the curves are given for hit ratios *(h)* with respect to various distances from AP$_2$ *(d$_2$)* in cases of $n_1 = n_2 = 6, 12, 18$ and 24.

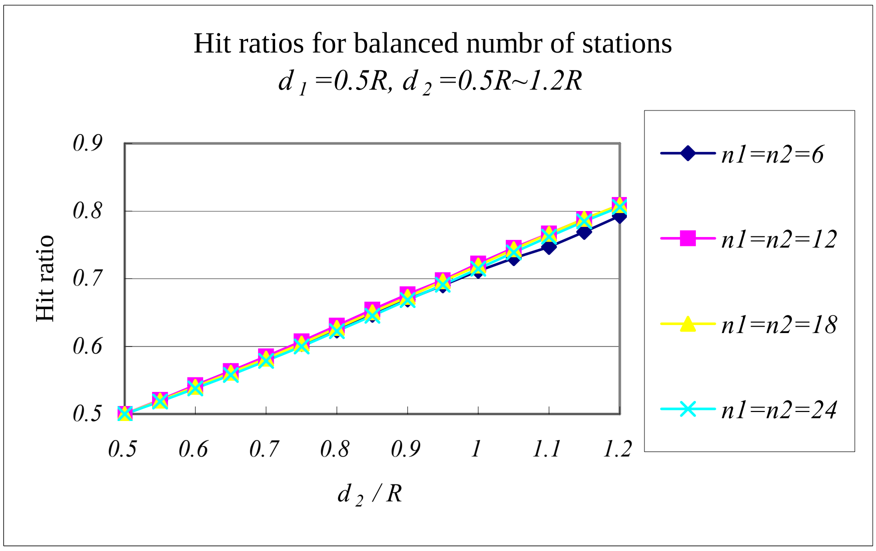


Fig. 16: Hit ratios vs distances in balanced number of stations.

Fig. 16 shows that the hit ratios increase almost linearly with the increase of distance between the mobile node and the competing AP (d_2), while the distance from the mobile node to the selected AP (d_1) is set fixed and the effect of imbalanced number of stations are ignored, i.e., $n_1 = n_2$. From observation of Fig. 16, it indicates that the number of stations, if they are evenly distributed, will be almost independent to the hit ratios.

In the second case, we investigate the affect of imbalanced number of stations on the hit ratios. The mobile node keeps its location at the place $d_1 = 0.5R$ and $d_2 = 0.9R$. Four cases of variation of number of stations in BSS$_1$, $n_1 = 6, 8, 10$ and 12 are individually input. For each case, the number of stations in BSS$_2$,

n_2, varies from *0.6* to *2* times of n_1 and the corresponding hit ratios are computed and presented in curves, as shown in Fig. 17.

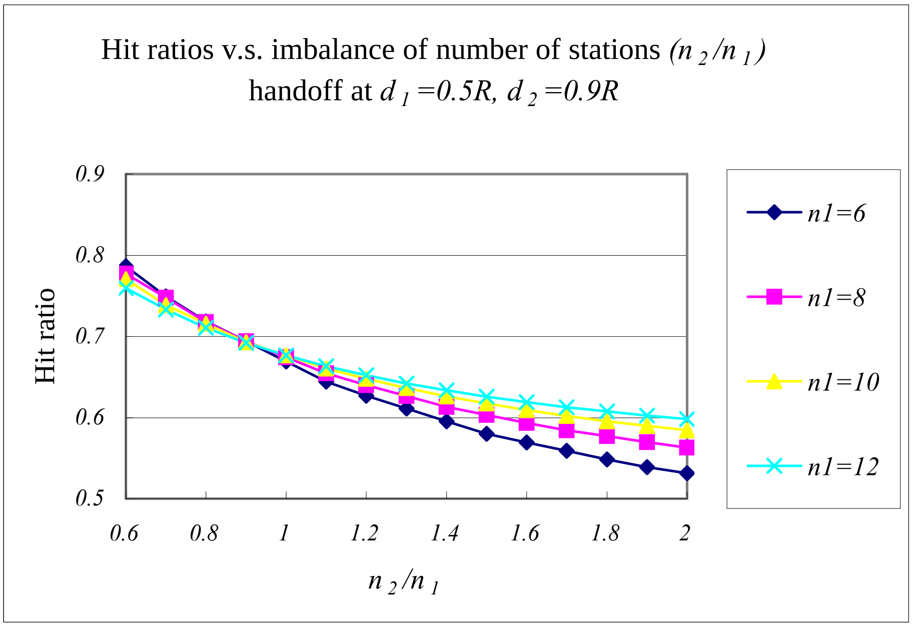


Fig. 17: Variation of hit ratios with imbalance of number of stations.

Observe the curves in Fig. 16, the imbalance of number of stations do affect the hit ratios. In the neutral number of stations, i.e., $n_2 = n_1$, the hit ratios for all of the four cases are about *0.68.* With the increase of imbalance, i.e., increasing n_2 / n_1, the hit ratios decrease with moderate slopes. As to the most imbalanced number of stations, i.e., case of $n_2 = 2n_1$, the hit ratios drop to about *0.55~0.6*. The hit ratio decreases by less than 20% due to the number of stations of competing BSS increases up to double. In addition, it is more sensitive to the disturbance of imbalance in number of stations in case of fewer stations in the selected BSS (BBS_1). From Fig. 17, in the case that most of stations are in selected BSS (i.e., n_1 = 12), the hit ratio drops to about 10% (from 0.68 to 0.6) with the increase of imbalance (n_2 / n_1) from 1.0 to 2.0. On the other hand, the hit ratio drops to about 20% (from 0.68 to 0.55) in case that there are fewest stations in the selected BSS (n_1 = 6).

2.5 Select best next AP instead of the nearest AP

The traditional handoff scheme that determines the next AP to handoff to by the RSS (receive signal strength) received from the responding AP in the probe phase of a layer-2 handoff. The strongest RSS is recognized as an indicator of the nearest AP, because shorter distance is subject to stronger signal strength. Same as RSS, the mechanisms presented before help a roaming node select the nearest AP by the link-layer information obtained in discrete scans. However, the nearest AP may not be the best next AP to handoff to. The available bandwidth that the next BSS can offer is one of the most concerned items especially for a mobile node in real time application, such as a WiFi VoIP connection. Besides, an AP that the mobile node is approaching to may be a better selection of the next AP than those from which the mobile node is moving away, even the latter ones are detected with stronger RSS.

Discrete scan schemes may be used to detect characteristics of a BSS by the collection of frames transmitted in a BSS in advance to a handoff. With information from sniffed frames, a mobile node may select the best next AP rather than the possible nearest AP with the indication of RSS in a traditional handoff scheme.

The available bandwidth and access delay of a BSS can be detected by the average NAV values in the frames collected in discrete scans. The relations between the NAV values and available bandwidth as well as access delay in a BSS are inferred to be linearly dependent in [6] with both mathematical analysis and simulation. A mobile node with discrete scan will be able to estimate the available bandwidths and access delays of the next-BSS candidates by NAVs collected in its sniffing periods and determine the best next AP when a handoff is triggered.

Furthermore, the trends of station counts along certain sniffing periods for a BSS may be used to indicate whether the sniffing node is approaching to an AP or not. For the AP that a mobile node is approaching to, the station counts shall be in an increasing trend because the mobile node can listen to more stations when it is closing to the AP. To select an approaching AP rather than the nearest AP may alleviate the frequent handoffs in the hot spot where the cell of a BSS may be relatively small.

The attributes that assist a mobile node to determine its next AP in handoffs may be integrated as an objective function with appropriate weights assigned to all of the indicators. With combinational considerations on receive signal strength, distances, approaches, as well as available bandwidth, the objective function for a mobile node to evaluate i^{th} BSS can be written as,

$$F_i = w_1(RSS)_i + w_2(STA\_count)_i + w_3(\Delta STA\_count)_i + w_4(NAV)_i \qquad (17)$$

2.6 Impact on service QoS

The most concerns on the discrete scan schemes may attribute to its impact on the service QoS in a pre-handoff period, since part of service time is taken out to sniff on other channels. During the sniffing periods, the service of working channel will be interrupted and this causes a certain level of access delay. Furthermore, because of absence from working channel, the frames that the current AP sends to the mobile node during a sniffing period will be lost. However, as described in chapter 1, the applications that require seamless handoffs are mostly of relative low bit rate as compared with that a wireless NIC can offer. Therefore, with proper design, a mobile node may use a large portion of idle time of wireless NIC for discrete scan; thus, to minimize the impacts caused by discrete scan. In this section, a VoIP connection which runs with bidirectional 64 kbps constant bit rate in an IEEE 802.11b wireless LAN of 11Mbps are taken as an example for the discussion on the impact on service QoS brought by discrete scan scheme.

Besides a sniffing period of 20 ms, the channel switch time of a wireless NIC contributes to disruption of service from working channel as a major part. In this project, the channel switch time is assumed as 5 ms as in [3]. To complete a sniffing period in discrete scan scheme, it needs to do twice for switching channels, one for leaving from the working channel and the other for returning back. Therefore, it brings an absence period of at least 30 ms from the working channel to execute once of discrete scan.

A disruption of 50 ms may be the upper bound for a seamless handoff to tolerate. The principle of the discrete scan is to decompose time for the probe phase of a layer-2 handoff into a pre-handoff period, such that no more than 50 ms disruptions are induced during pre-handoff and handoff procedures.

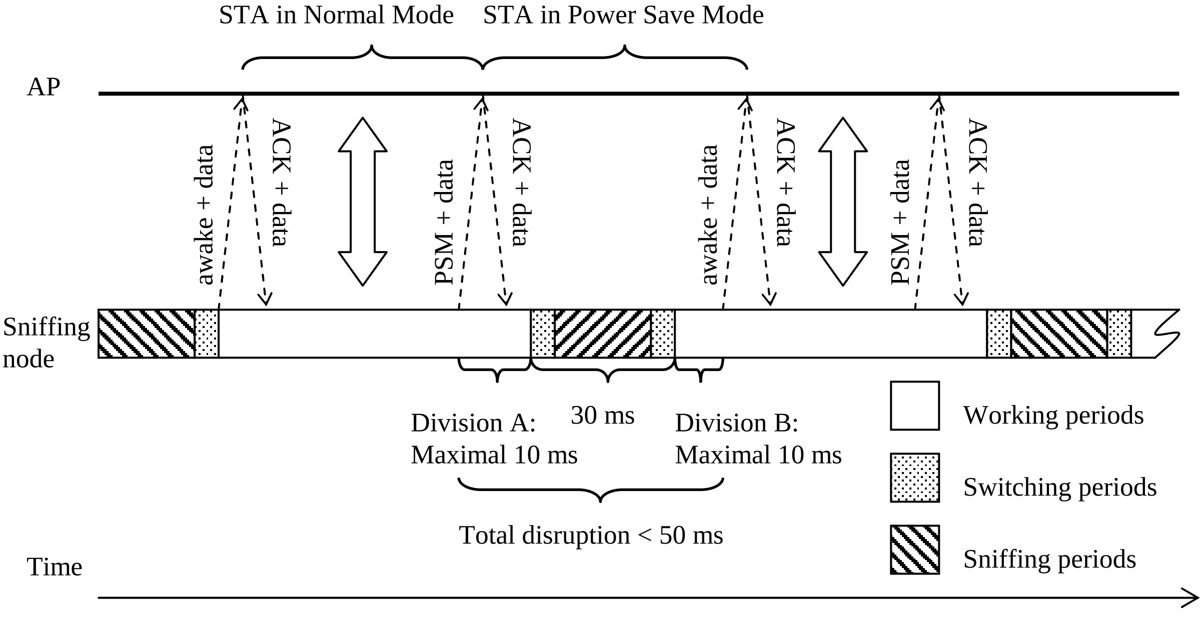


Fig. 18: Total disruptions induced with a discrete scan.

Nowadays, the QoS issue for real time applications in IEEE 802.11 wireless LAN is still widely discussed. In most of infrastructure wireless LAN environment, downlink traffics from AP to all the stations is much heavier than the uplink traffic from a single station to the AP. However, the transmission opportunity of AP is same as that of a single station, thus the downlink flow shares the bandwidth with those uplink flows from all stations. In other words, the shared bandwidth of the downlink transmission from AP to arbitrary one of stations will be only $1/n$ of the bandwidth of the uplink transmission to AP, where n denotes number of stations in a BSS. Because of the asymmetric nature between uplink and downlink transmission, a real time application with symmetric bidirectional transmission may suffer from severely degraded QoS of downlink flow when the number of stations in a BSS increases.

To address the QoS problems caused by asymmetric transmission opportunity for a specific station with symmetric bidirectional traffics, piggyback schemes [7] in IEEE 802.11 standard suppose to forward the downlink traffic as an attachment to a positive ACK frames from AP to stations. With piggyback schemes, the QoS issue of asymmetric transmission opportunity is therefore eliminated. The QoS provisioning to symmetric bidirectional service then is considered as the QoS to the uplink flow of the concerned station.

As illustrated in Fig. 18, the total disruption caused by discrete scans should be computed from the transmission of the last frames before wireless NIC switches to a sniffing channel until the transmission of the first frame after wireless NIC returns back to working channel. The duration for a wireless NIC absent from working channel is 30 ms for each sniffing period and the total disruption has to be kept less than 50 ms, therefore, the two sentinel frames, defined as the last frame before channel switch and the first frame after the return, have to be transmitted within the periods of 20 ms in total.

The rest of 20 ms for sentinel frames is divided into division A and B (shown in Fig. 18). A mobile node converses a disruption caused by a sniffing period less than 50 ms, if it transmits the first sentinel frame in division A and the second sentinel frame in division B. In the case the first sentinel frame fails to be transmitted with the division A period, the piece of sniffing period should be slipped until next cycle. As the first sentinel frame has been sent, the total disruption brought by the sniffing period is equal to 50 ms + (time to the 2nd sentinel frame – division B) – time to the 1st sentinel frame. To keep the disruption less than 50 ms in all case, a mobile node should promise to send the second sentinel frame by the end of division B.

In the simplest design, each sentinel frame for a sniffing period evenly shares 10 ms as its maximal allowable time for the mobile node to transmit a sentinel frame successfully. Thus we have division A = division B = 10 ms. Therefore, the capability that a mobile node guarantees to send one frame successful in 10 ms decides whether the disruptions brought by a discrete scan can be conserved within 50 ms. In this section, the analytical model of DCF wireless environment presented before is employed to estimate the probability for a mobile node in IEEE 802.11b wireless LAN to transmit one frame successfully within 10 ms.

With known transmitting probability in a slot time, τ, given in (2) and the average length of a slot time, T_{av}, given in (8), the average number of frames that the mobile node could transmit within 10 ms can be derived as,

$$m'' = \tau * 10ms / T_{av} \qquad (18)$$

Taking the probability that a transmitted packet encounters a collision, p, given in (5), the probability that at least one of m'' frames transmitted without a collision is estimated by,

$$p_{sentinel} = 1 - p^{m''} \qquad (19)$$

Numerical results of (19) with the parameters listed in Table 1 for BSSs of 11 Mbps as well as 54 Mbps are given in Fig. 19. When number of stations increases up to 6, the probability for a station to transmit a frame without collisions within 10 ms drops to lower than 80% in an IEEE 802.11b 11Mbps wireless LAN, therefore, more than 20% of sniffing periods will be slipped. It is inferred about 20% of disruptions brought by sniffing periods to be longer than 50 ms when the BSS supports more than 6 stations simultaneously. Figure 19 shows that even bit rate of the BSS increases up to 54 Mbps, stations that the BSS can support based on the same criteria increases to only 12. Note that, as shown in Fig. 16 and 17, the station count mechanism proposed before assumes a number of stations greater 6 to promise creditable selections. It consequently concludes that the discrete scan with station count mechanism proposed before will not work together very well in the DCF environments. However, instead of DCF, an IEEE 802.11e EDCA wireless environment grants VoIP traffic with high priority of transmission and promise it to send one frame within 10 ms at a much higher probability even with more than 10 stations in the BSS.

How discrete scan and station count mechanism proposed in this project can work together in an IEEE 802.11e ECDA wireless environment will be elaborated with results of ns-2 simulation. To set a proper length for a sniff period, the performance to hit the desired nearest AP as well as probabilities that the disruptions for sniffing periods converge within 50 ms will be demonstrated.

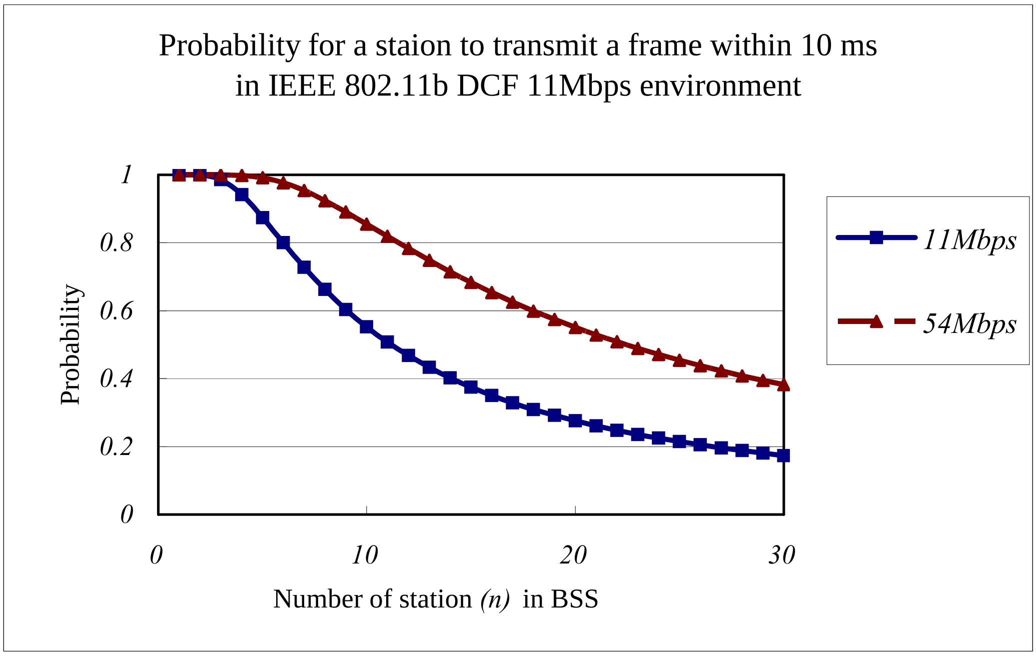


Fig. 19: Probability to transmit a frame in 10 ms vs number of stations in a BSS.

2.7 Performance evaluation

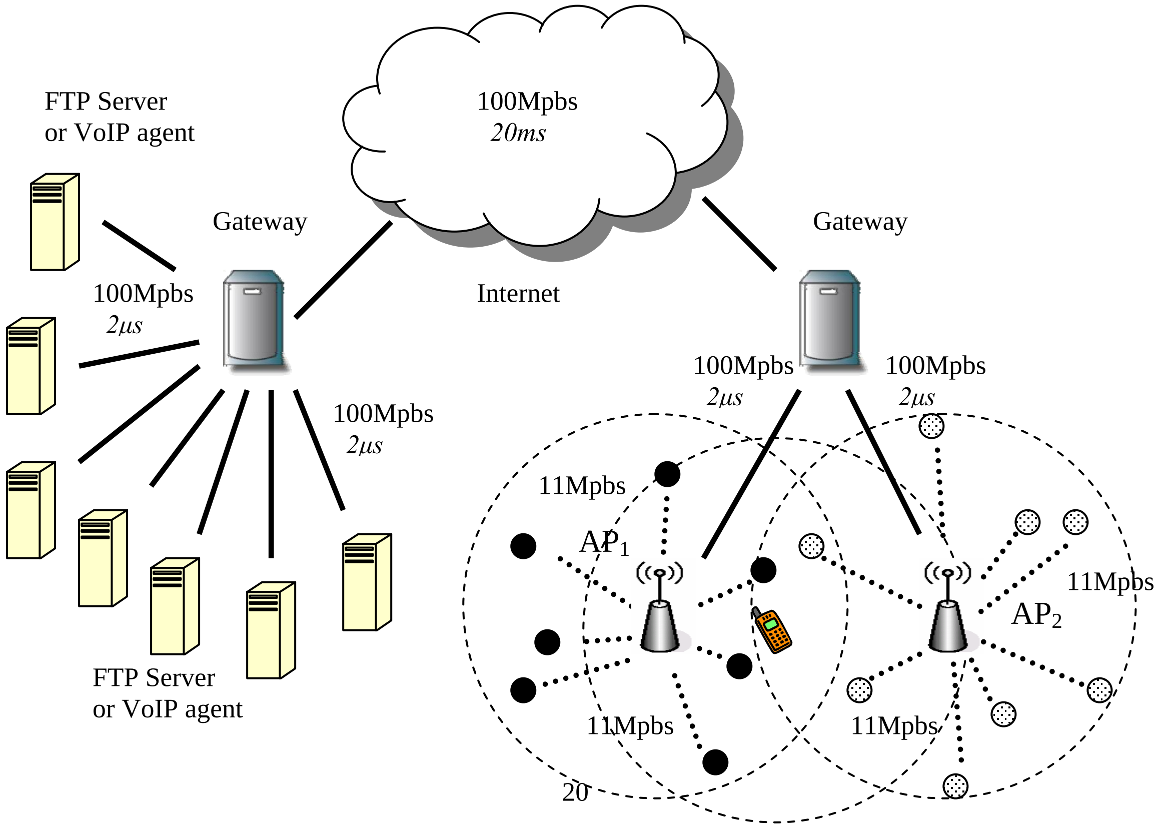


Fig. 20: Simulation configuration.

We evaluate the performance for discrete scan schemes in IEEE 802.11e EDCA wireless LAN environment by means of simulations. We discuss the major concern, the impact of service QoS, while a mobile node applies discrete scan scheme in EDCA wireless LAN. Attributing to the high transmission priority of VoIP traffic, the impact on QoS caused by discrete scan schemes in EDCA environment is shown within the tolerable limits according to the simulation results.

Simulation Environment

We use NS-2 (version 2.28) tool [8] with 802.11e tkn EDCA module [9], we neglect high-level management functionality such as beacon frames, association and authentication frame exchanges. In all scenarios, the network topology of the simulations is shown in Fig. 20. Each of wireless stations either runs bidirectional VoIP traffic or TCP traffic with its corresponding wired station playing as either a VoIP device or FTP server. VoIP traffic, with format of G.711 codec, 160 bytes payload and 20ms intervals are used as real-time traffic, and FTP traffic of 1,500 bytes payload are used to simulate the best effort traffic. Table 2 shows the default parameter values of EDCA in the simulations. Besides, 10% of wireless packet error rate is set to simulate the transmission in the real environment. Furthermore, for the sake of distributed arrival time of VoIP packets, each of VoIP traffic is initiated at a randomly selected time from sec 5th to 6th sec. The observed scan periods for each channel are taken from 10th sec to 20th second, so that the traffics in BSSs are assumed in steady state. The simulation ends at 25th second.

With various random seeds for each run, uniformly distributed random function which is built-in in ns-2 programs generates the locations of the mobile nodes in both BSSs. The normal combination of traffics is assumed that 70% of nodes run in TCP traffic and 30% of nodes play as VoIP phones. However, BSSs of all TCP nodes and all VoIP nodes are running in order to discuss the affects by various traffics in a BSS.

AC	PF	AIFS	CW_MIN	CW_MAX	TXOP Limit (s)
voice	2	2	7	15	0.003008
best effort	2	3	31	1023	0

Table 2: Default values of parameters for original EDCA.

Station Count Mechanism in EDCA Environment

As discussed before, by properly selecting a sniffing time may eliminate affects on the hit ratio in selecting the nearest AP caused by imbalance of number of stations between the next-BSS candidates. The number of transient stations (denoted by *r*) is defined as the average number of stations that transmit at least once within a sniffing period (denoted by T_{sniff}) in a BSS. A proper sniffing time shall meet two conditions: first, make the number of transient stations varies as insensitively as possible with the variation of number of station in a BSS. Usually the less the sniff time is, the better the attribute will be. Second, the sniff time shall be long enough for a mobile node to discover the next-AP candidates as well as to collect enough frames to infer a creditable selection result.

In Fig. 21, it shows the curves for transient number of stations v.s. various numbers of stations for several given values of sniffing periods in an EDCA BSS of 70% TCP and 30% VoIP nodes. The curves in Fig. 21 show that number of transient stations is fairly insensitive to number of stations as the number of stations in a BSS is large (e.g. $n > 12$). When the number of stations in a BSS is small (e.g. $n < 6$), the number of transient stations increases at a rate about a half of increasing rate of number of stations. The selection results are still disturbed by the imbalance of number of stations, especially when number of stations is small in both EDCA and DCF wireless environment. However, the effect is reduced to at least half of the difference of number of stations by setting a sniff period to 15 ~ 20 ms.

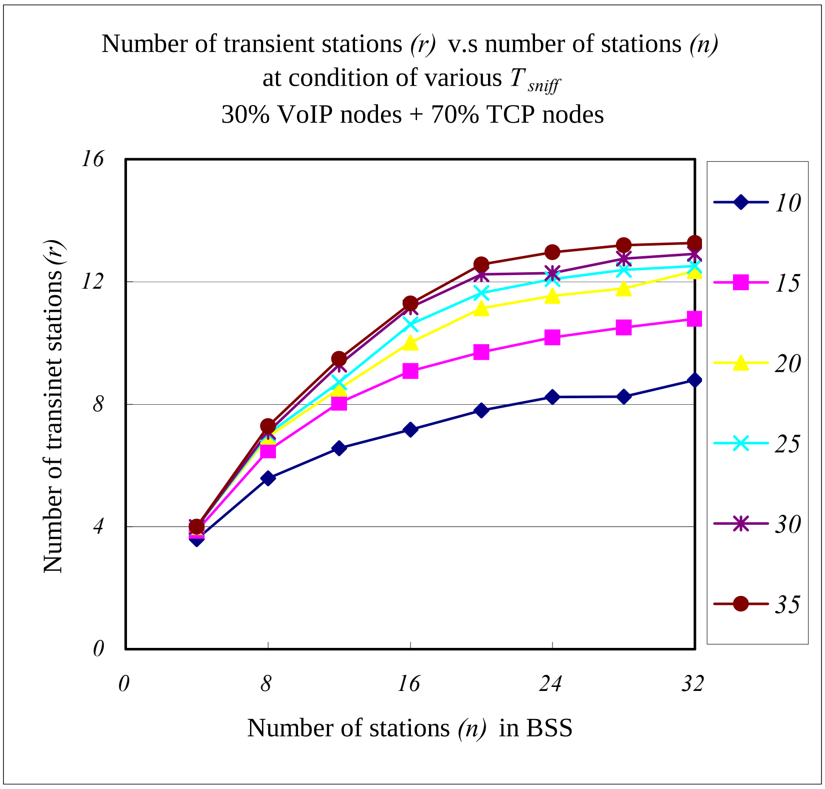


Fig. 21: Number of transient stations v.s. number of stations in EDCA of 70% TCP and 30% VoIP nodes.

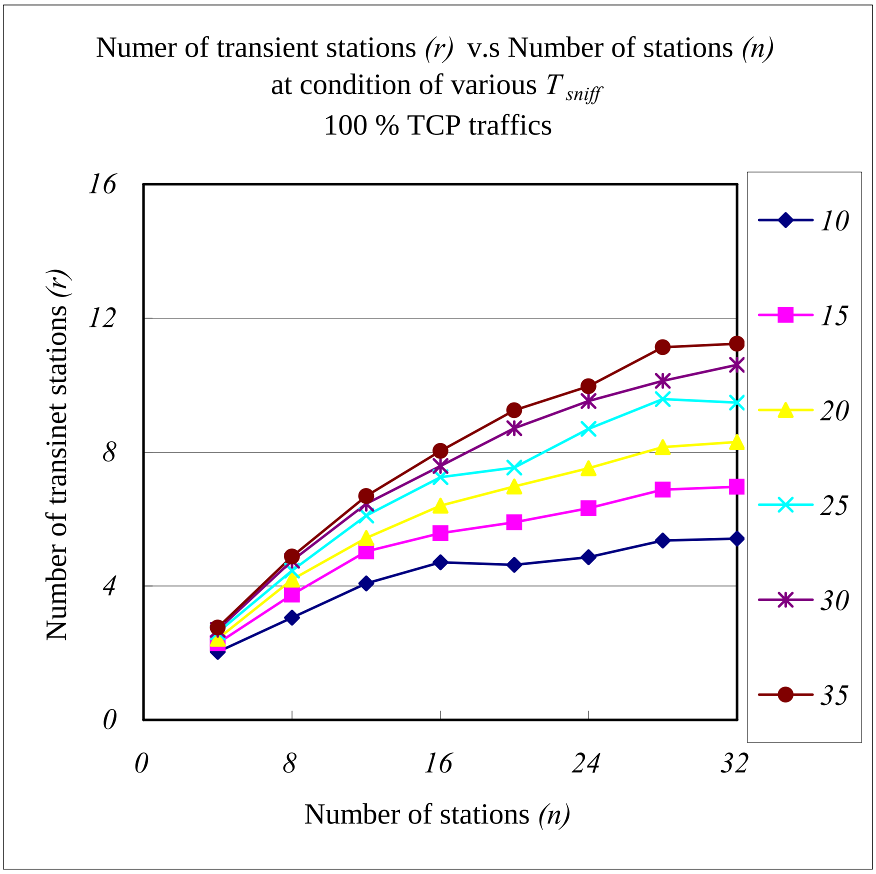


Fig. 22: Number of transient stations v.s. number of stations in EDCA with 100% TCP nodes.

Figure 22 shows the curves with the same condition as those in Fig. 21 except that EDCA BSS features 100% of TCP nodes. With mostly the same trends as that in Fig. 21, the curves in Fig. 22 shows less steep slope and apparently lower values of number of transient stations than those in Fig. 21. It is inferred that most VoIP nodes are discovered and contribute to number of transient stations because of the attributes of high transmission priority and short packet length of VoIP traffic. Consequently, the station count mechanism tends to select a BSS with more VoIP nodes as the next AP to handoff to if the other control factors, such as distances and number of stations are neutral between BSSs.

Performance of Station Count Mechanism in EDCA Environment

Figure 23 shows that the hit ratios with respect to various distances (d_2 = 0.5R ~ 1.1R) from the competing APs. The selected AP locates at a fixed distance (d_1 = 0.5R) from the mobile node. The number of stations in both BSSs are set to the same set of numbers ($n_1 = n_2$ = 6, 12, 18, 24) to minimize the effects caused by imbalanced number of stations. Both next-BSS candidates work as EDCA wireless LANs consisting of 70% TCP and 30% VoIP nodes.

As the same situation as those in DCF environment (shown in Fig. 16), the hit ratios increase linearly with the increase of the distance between the mobile node and the competing APs, d_2, while d_1 is fixed and number of stations are evenly distributed. It is inferred that the distances between a mobile node and next-AP candidates affect the hit ratios to select the nearest AP regardless the type of wireless LANs, with the assumption of evenly distributed stations. Consequently, it concludes that the station count mechanism works under EDCA as well as under DCF.

Figure 24 shows the hit ratios with the same as condition in Fig. 23 except that the EDCA BSS consists of 100% TCP nodes. Compared with that in Fig. 23 and Fig. 16, it is inferred again that the distance dominates the hit ratios linearly. However, the deviation among curves of numbers of stations (n_1 and n_2) is apparently larger than that in Fig. 24. As described in previous section, a mobile node will discover most of the VoIP nodes in its coverage; therefore, these VoIP nodes contribute to the station count. The curves in Fig. 24 display a trend much similar to those in Fig. 16 because EDCA is reduced to DCF if only best effort access category (AC) exist in an EDCA BSS.

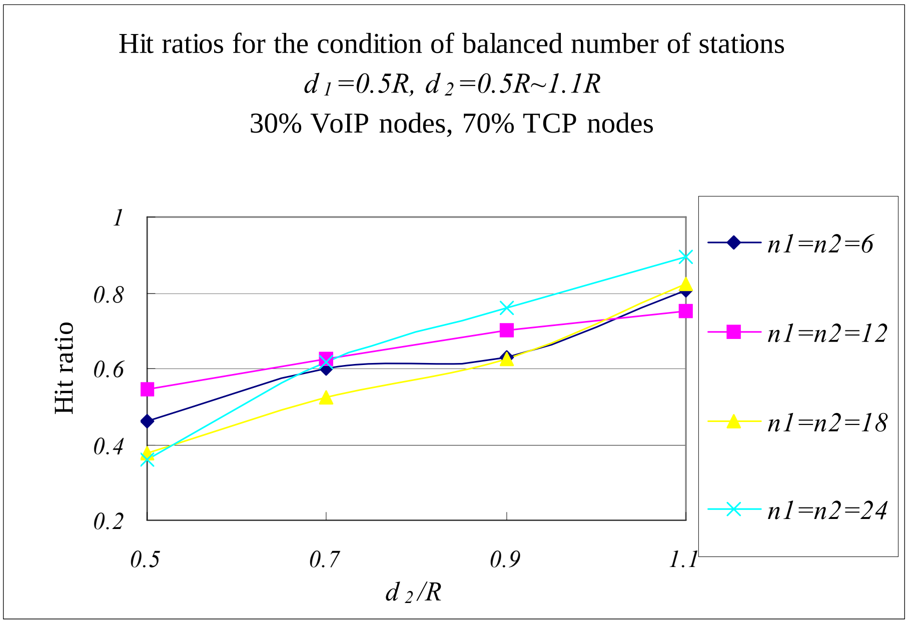


Fig. 23: Hit ratios v.s. distances in balanced number of stations in EDCA WLAN of 30% VoIP nodes and 70% TCP nodes.

Furthermore, with existence of VoIP nodes in EDCA, small portion (e.g. 30%) of VoIP nodes dominate the selection of the next AP. The station count mechanism may select a BSS of more VoIP nodes rather than the nearest one.

Figure 25 shows the effects on hit ratios caused imbalance number of stations in an EDCA BSS of 70% TCP nodes and 30% VoIP nodes. Same as those in Fig. 17, the mobile node locates at the place $d_1 = 0.5R$ and $d_2 = 0.9R$, the number of stations of BSS_1, $n_1 = 6, 8, 10$ and 12, and the imbalance of number of stations of BSS_2, n_2/n_1 varying from 0.6 to 2.

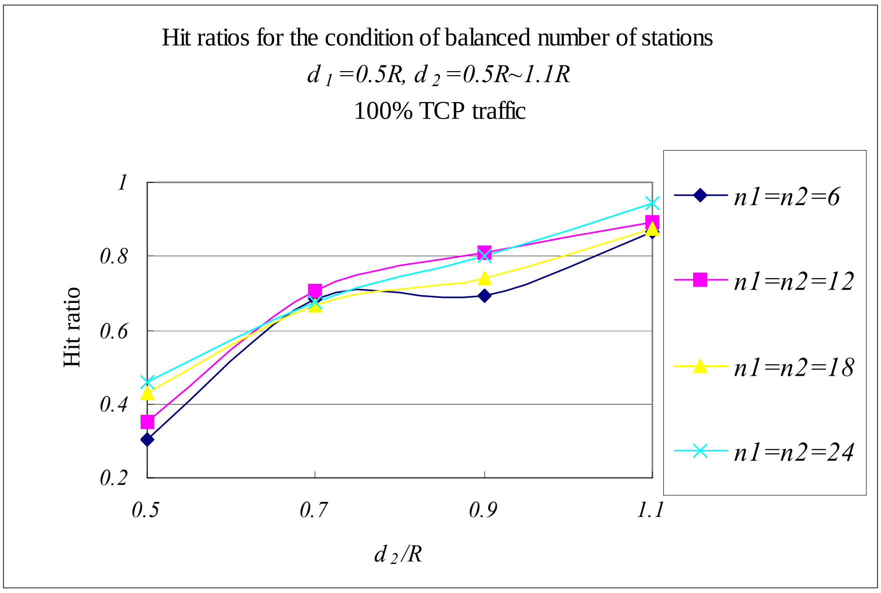


Fig. 24: Hit ratios v.s. distances in balanced number of stations in EDCA WLAN of 100% TCP traffic.

Figure 26 shows the hit ratios of the same settings in Fig. 25 except that the EDCA has only TCP nodes. Figure 25 and Fig. 26 show the interference of imbalanced number of stations in the next EDCA BSSs candidates on the hit ratios of station count mechanism are of the same averages and trends, except that the curves in Fig. 25 are of larger deviation from their average values. As interpreted above, the existence of VoIP nodes contribute the deviation to the curves in Fig. 25, because few VoIP nodes dominate the selection results.

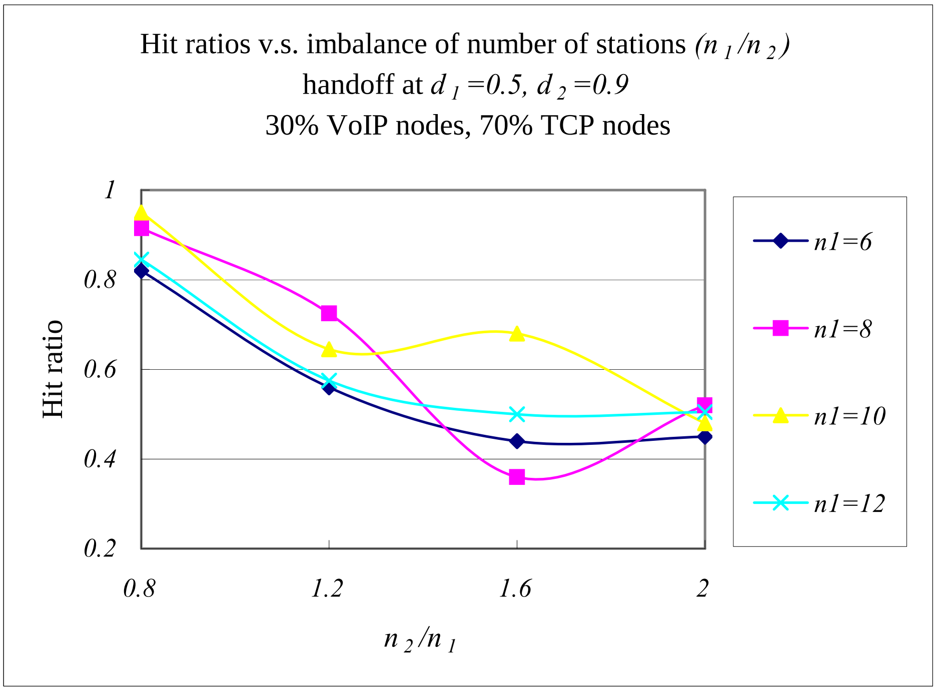


Fig. 25: Hit ratios v.s. imbalanced number of stations

in EDCA of 30% VoIP nodes and 70% TCP nodes.

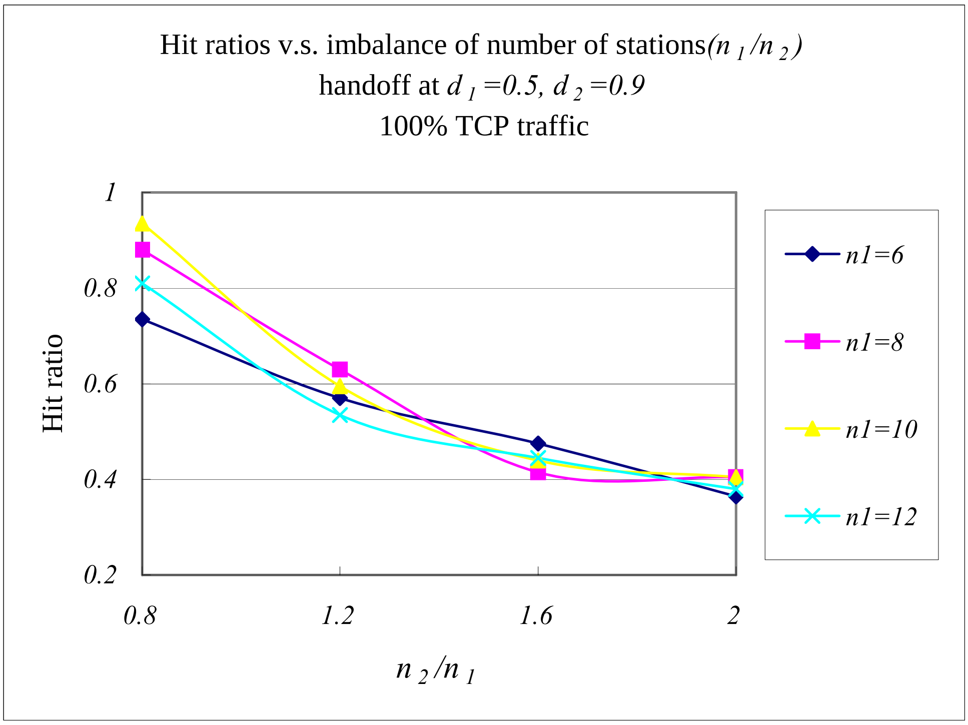


Fig. 26: Hit ratios v.s. imbalance of number of stations in EDCA of 100% TCP traffic.

With the discussion above, the existence of VoIP nodes may contribute a negative factor for station count mechanism to select the nearest AP. Generally speaking, VoIP packets are generated in an interval of 20 ms and transmitted within bounded jitter to meet real-time requirement. Since a sniffing period is suggested to be as long as 20 ms, it implies that all VoIP nodes transmit a frame each sniffing period. Therefore, the number of VoIP nodes is included in the number of transient stations in an EDCA BSS, then, it causes station count mechanism intending to select a BSS of more VoIP nodes rather than the nearest one.

It is still not yet to comment as a good or bad feature for station count mechanism to select an AP of more VoIP nodes in its BSS rather than the nearest one. However, the feature can be designed as an option because of the simplicity to filter out the frames from VoIP nodes by the specific characteristic of its packets. As the frames for VoIP traffic are ignored in the sniffing periods, the performance of the station count mechanism in EDCA environment reforms to be similar to that in DCF and with less deviation from the mean values, as shown in Fig. 24 and Fig. 26.

Impact on Service QoS in EDCA Environment

Referring to the discussion before and the scheme shown in Fig. 18, to conserve the seamless requirement and protect services from a disruption of more than 50ms, a VoIP node with discrete scan scheme shall ensure to transmit one frame successfully within 10 ms before and after a sniffing period. In an EDCA wireless LAN, VoIP traffic is granted with the highest transmission priority, therefore, the contentions for a VoIP flow in EDCA wireless LAN to get medium are much less intense than those in DCF one. In Fig. 27, it shows the probabilities for a VoIP node to send at least one frame in 10 ms with respect to the number of stations in an EDCA wireless LAN with 11Mbps rates. Three cases for further discussion are assumed: first, 10% of transmission error in the wireless environment and 100% of stations in a BSS are VoIP nodes; second, 10% of transmission error in the wireless environment and normal traffic combination (30% VoIP nodes and 70% TCP nodes) for stations in a BSS; third, in the error free wireless environment and 100% of VoIP nodes for all stations in a BSS.

The curves in Fig. 27 display that the probabilities for a VoIP node to send a frame in 10 ms in the EDCA environment are more dependent on the air transmission error rather than on the number of stations in its BSS. In worst condition such that more than 20 VoIP running simultaneously in an EDCA BSS with 10% transmission error, the success probability to send a frame in 10 ms is still kept about 85%. Therefore, the disruption caused by a sniffing period for discrete scan will be maintained within 50 ms with a probability more than 85%. As compared with those for DCF environment, it concludes that the discrete scan schemes proposed in this project are much applicable for a VoIP node in EDCA wireless environment than in DCF one.

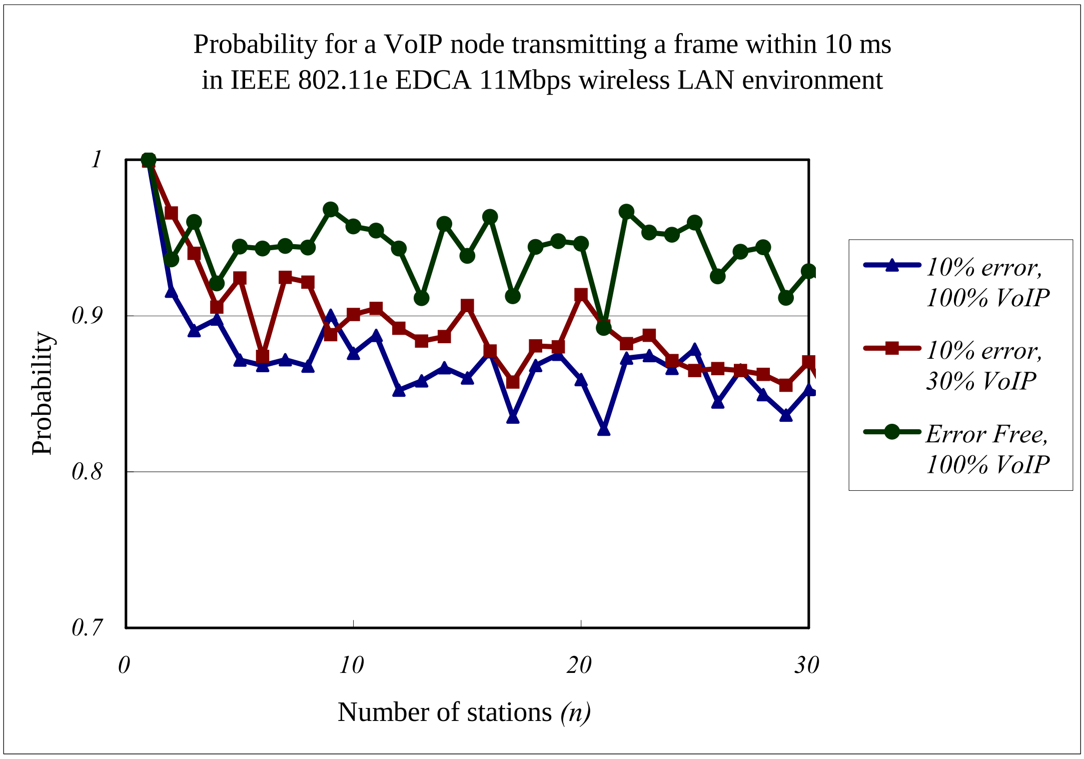


Fig. 27: Probability to send a frame in 10 ms v.s. number of stations
in 11 Mbps EDCA wireless LAN.

2.8 Section conclusion

Compared with neighbor graphs, discrete scan scheme provides a set of next-AP candidates as a subset of neighboring APs. With the help of proposed auxiliary mechanisms, discrete scan selects an appropriate next AP for a mobile node to handoff to, taking the place of the function of a neighbor graphs with aids of topological information by GPS system proposed in [4].

As a cost, discrete scan scheme contributes an impact to service QoS received by the mobile node in a “pre-handoff” period that is defined as the duration from initiating the discrete scan scheme to the moment when a handoff is initiated. However, taking the advantages of high transmission priority and relative low bit rate for a VoIP connection in EDCA wireless LANs, the degradation may be limited to a degree that the users may tolerate or even may ignore. Estimation on the QoS degradation with analytical approach and simulation results presented in this and next chapter shows that the disruptions caused by discrete scan are bounded to 50 ms based on certain practical assumption.

3. Two-stage Authentication

3.1 AAA/Mobile IP

We propose a two-stage authentication scheme such as transient authentication and user (re) authentication mechanism. We use AAA/Mobile IP to authenticate a MN and get the certificate from MAP during registration simultaneously. This certificate signed by MAP can be used in the same MAP domain to get the different local certificate from different access routers for passing the filtering rule of MN. Figure 23 shows the authentication message flow in AAA/Mobile IP as shown in Figure 28. A mobile node sends its NAI in registration request and AAAF forwards registration message with authentication information to AAAH. Then, AAAH generates 3 keys to communicate with the MN and FA and sends them in registration reply message. After registration process, MN derives 2 keys such as MN-FA and MN-HA keys to encrypt data on the Internet. We use this characteristic in our Hierarchical Mobile IP with Fast handover experiment. If mobile node first enters into the new Hierarchical Mobile IP domain, it will perform AAA/Mobile IP initial registration and regional registration. When MAP receives the initial registration reply from HA, it checks the information in this message and adds an extra certificate signed by MAP called CAmap on registration reply message to the MN. Then, the access routers under MAP can authenticate the MN through CAmap and send out a new local certificate signed by the access router.

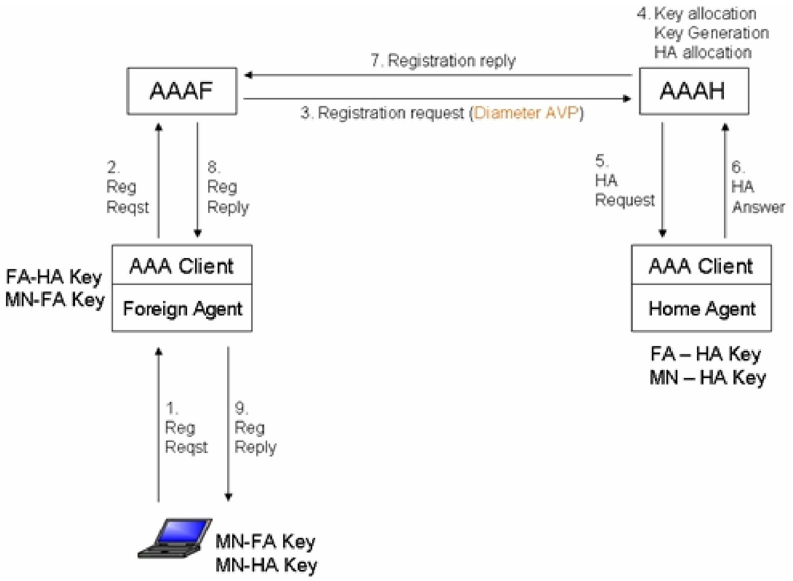


Fig. 28: AAA/Mobile IP Initial Registration.

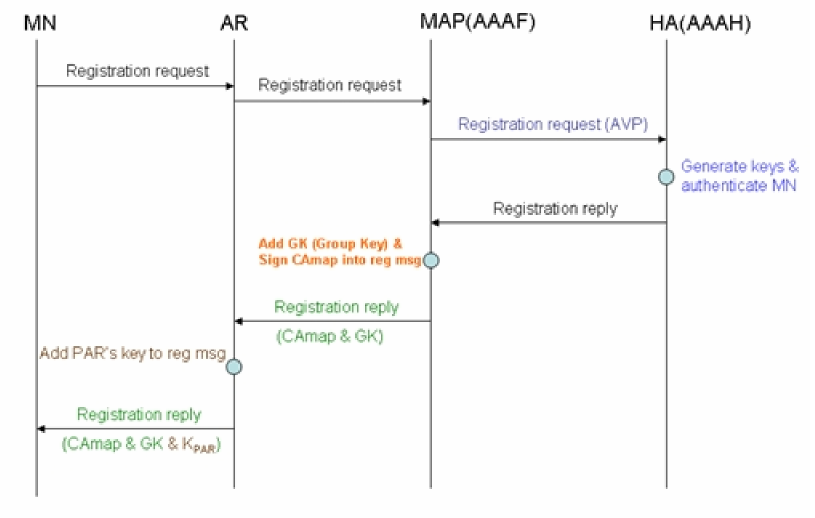


Fig. 29: CAmap & GK & Key_{PAR} Distribution Flow.

Figure 29 shows the GK and CAmap distribution during AAA/Mobile initial registration. After initial registration, MN derives two keys from HA (AAAH), one key (GK) from MAP and one key (K_{PAR})

from current access router. The group key is used to encrypt packets for transmission in the MAP domain. The other key distributed by an access router is used to encrypt packets between mobile node and access router. This is the last mile key. A mobile node will use these two keys to transmit some information to new access router through current access router based on fast handover control messages, this action is called transient authentication. We describe this content in later sections.

3.2 Two-Stage authentication scheme

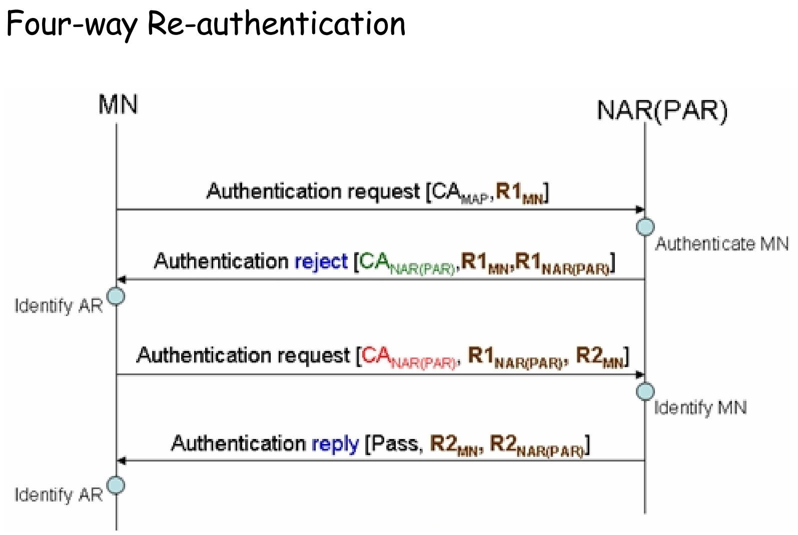

Four-way Re-authentication
MN
NAR(PAR)
Authentication request [CA_{MAP},$R1_{MN}$]
Authenticate MN
Authentication reject [$CA_{NAR(PAR)}$,$R1_{MN}$,$R1_{NAR(PAR)}$]
Identify AR
Authentication request [$CA_{NAR(PAR)}$, $R1_{NAR(PAR)}$, $R2_{MN}$]
Identify MN
Authentication reply [Pass, $R2_{MN}$, $R2_{NAR(PAR)}$]
Identify AR

Fig. 30: 4-way Re-authentication.

There are several approaches that filter packets toward mobile node as follows. MAC filtering is a common method to drop the illegal MAC address packets of mobile node. A switch or router would construct a table to record which node passes the authentication. Currently, the most popular user authentication is the web-based approach. A mobile user just inputs the correct password to the web interface and sends it to the authentication server to open the access right for himself/herself. Then, the mobile user can get the access right to the Internet. The re authentication we propose is to use certificate to derive the access right to the Internet. Also, each access router can identify the certificate of MAP and sign it for the mobile user. Access routers use their own certificate to authenticate the mobile user and to pass the packets over to the mobile node.

We define a 4-way re-authentication signaling to simulate user authentication process similar to AAA operations. If a mobile node doesn't have a certificate of access router, it can send certificate of MAP to derive a certificate of the access router, which can authenticate mobile node by identifying CAmap. An access router signs a local certificate of its own to a mobile node for accessing the Internet under its coverage. When an access router receives the correct local certificate, it will stop dropping the packets destined to the mobile node, shown as the third message in Fig. 30. Also, we use challenge/response to identify each other against the reply attack during transmission. Currently, there are two authentication modes, unilateral client authentication and mutual authentication. The former is that the client digitally signs a challenge from the server, thus authenticating the client to the server. Mutual authentication is that the client digitally signs a challenge produced by the server and the server digitally signs a challenge produced by the client. Thus both the client and server authenticate each other. So, we adopt the mutual authentication mode, as Fig. 30 shows. First, a mobile node sends its $R1_{MN}$ random number to the access router. Then, the access router replies with $R1_{NAR\ (PAR)}$ and $R1_{MN}$ random numbers. The MN receives these random numbers and check whether they are correct. Moreover, the mobile node also replies with $R1_{NAR\ (PAR)}$ and a new random number $R2_{MN}$. The access router checks this random number to identify whether the message is freshness. After first three steps, they can identify each other and complete the authentication process. Through this 4-way re (user) authentication, the access router records the information of the mobile node and passes the packets to it. Packets destined to the mobile node are dropped by access router if the destination address is not be authenticated. An access router needs to maintain a table for authenticated users and a timer for refreshing the entry if it expires. The access router decides to drop or forward this packet according to the existence of destination address of the packet in the table. We can see that our simulation 4-way signaling is similar to IEEE 802.1x re-authentication messages from Fig. 31. So, we use this method to simulate the authentication delay during handoff.

Co-work with IEEE 802.1x Re-authentication

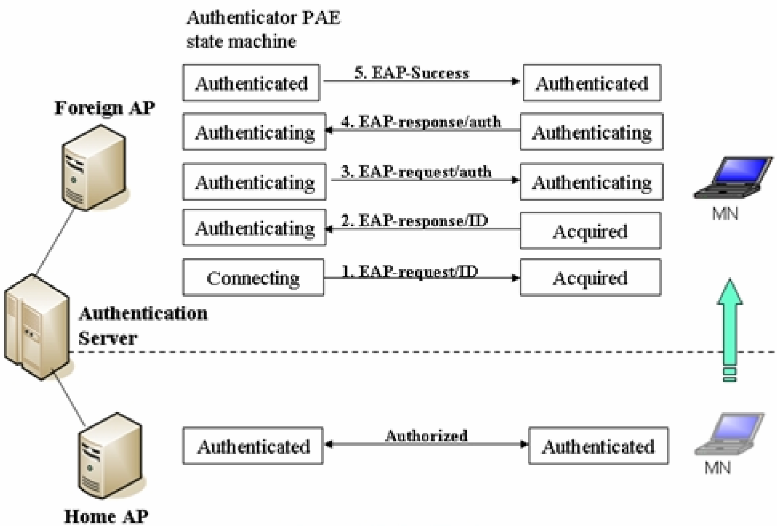


Fig. 31: Re-authentication State Transition in Roaming.

User authentication can be replaced by other user authentication protocols such as IEEE 802.1x. Transient authentication also can co-work with IEEE 802.1x user authentication scheme. So, the two-stage authentication is transient and IEEE 802.1x processes. If IEEE 802.1x supports transient authentication, it needs to open the controlled port for the mobile node temporarily to receive data according to the authentication table. If the lifetime for transient authentication expires, the authenticator will check if the IEEE 802.1x state is authenticated. If IEEE 802.1x state is not authenticated and lifetime for transient authentication expires, the authenticator will close the controlled port until IEEE 802.1x state is authenticated. Fig. 31 shows the IEEE 802.1x state transition in roaming [11].

3.3 Transient authentication

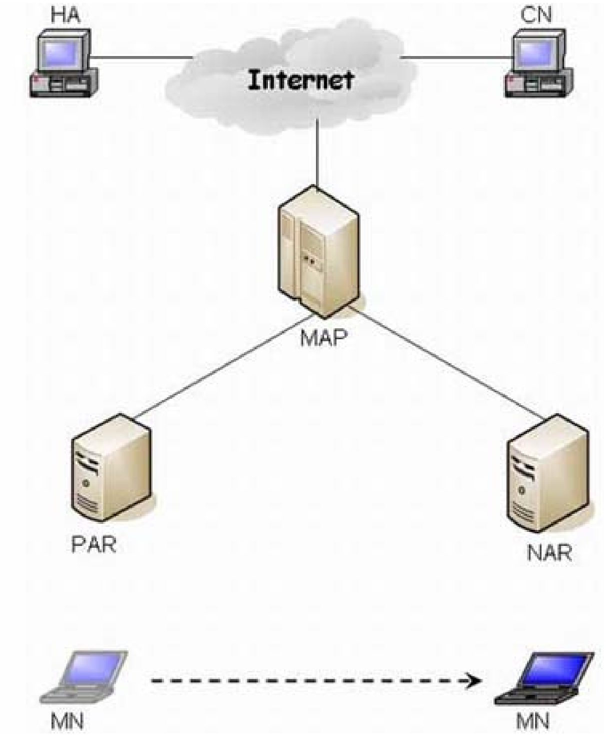


Fig. 32: Scenario for Handover.

There are several approaches to reduce handoff time, as discussed in Chapter 2. We choose the Hierarchical Mobile IP with fast handover as our baseline architecture, because it features the lowest handoff time among all available approaches. We can add an extra authentication signaling to increase a little handoff time; then we can also provide authentication function. Otherwise, it may cause a significant delay due to the original large handoff time. The scenario for handover is illustrated in Fig. 32. The mobile node first enters into the hierarchical structure, and it performs Mobile IP registration, Regional registration and re authentication. When the mobile node roams to NAR, it performs regional registration and re (user) authentication. A MN needs to spend some time doing authentication before receiving packets. Whether the time is long or short depends on the complexity of authentication. If the mobile device often roams between access routers, we should try to reduce the authentication time by transient authentication. We observe the control message of fast handover protocol and try to modify it for providing a transient authentication function. Our main objective is to derive a temporary access right of new domain through fast handover signaling. Also, we try to get a new local certificate of NAR through fast handover control messages. So, the new access router not only pre-registers new care-of address of mobile node but also temporarily authenticates mobile node through these control signaling. It can reduce handoff and authentication delay through this method.

Figure 33 shows a scenario of transient authentication between access points. We adopt the similar concept to perform our transient authentication in fast handover signaling. From Fig. 33, there is a pre-existing secure connection between oAP and nAP. This pre-existing connection is to ensure transmission between them safe.

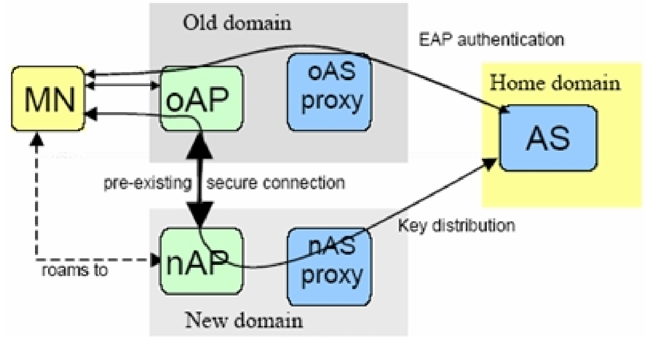


Fig. 33: Transient Authentication Scenario between APs.

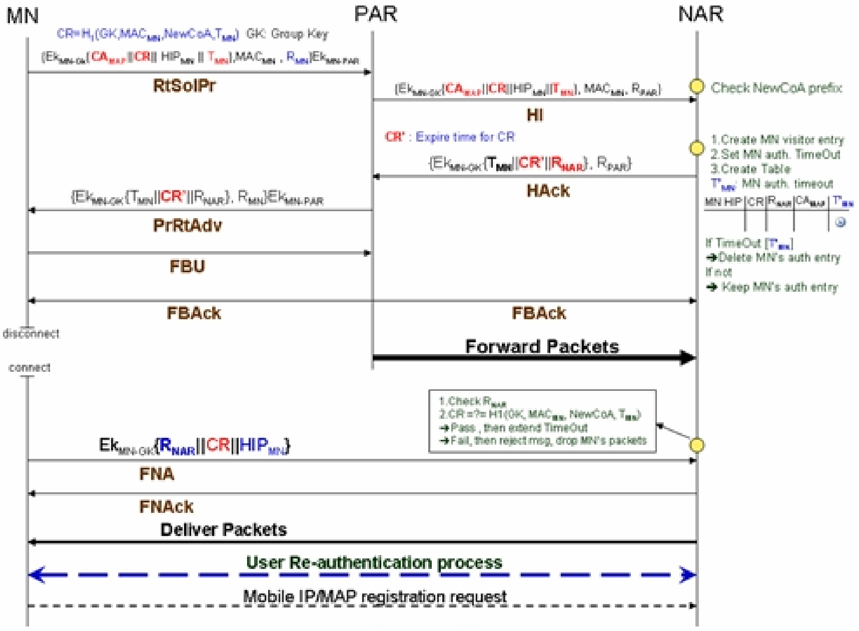

MN
PAR
NAR
$CR=H_1(GK,MAC_{MN},NewCoA,T_{MN})$ GK: Group Key
$\{Ek_{MN-GK}\{CA_{MAP}\|CR\| HIP_{MN} \| T_{MN}\},MAC_{MN}, R_{MN}\}Ek_{MN-PAR}$
RtSolPr
$\{Ek_{MN-GK}\{CA_{MAP}\|CR\|HIP_{MN}\|T_{MN}\}, MAC_{MN}, R_{PAR}\}$
Check NewCoA prefix
HI
CR' : Expire time for CR
1.Create MN visitor entry
2.Set MN auth. TimeOut
3.Create Table
T'_{MN}: MN auth. timeout
$\{Ek_{MN-GK}\{T_{MN}\|CR'\|R_{NAR}\}, R_{PAR}\}$
HAck
MN HIP | CR | R_{NAR} | CA_{MAP} | T'_{MN}
$\{Ek_{MN-GK}\{T_{MN}\|CR'\|R_{NAR}\}, R_{MN}\}Ek_{MN-PAR}$
PrRtAdv
If TimeOut [T'_{MN}]
➔Delete MN's auth entry
If not
➔ Keep MN's auth entry
FBU
FBAck
FBAck
disconnect
Forward Packets
connect
1.Check R_{NAR}
2.CR =?= H1(GK, MAC_{MN}, NewCoA, T_{MN})
➔Pass , then extend TimeOut
➔Fail, then reject msg, drop MN's packets
$Ek_{MN-GK}\{R_{NAR}\|CR\|HIP_{MN}\}$
FNA
FNAck
Deliver Packets
User Re-authentication process
Mobile IP/MAP registration request

Fig. 34: Transient Authentication Messages & Fast Handover Signaling.

Our architecture is similar to Fig. 33. A mobile node in old domain performs AAA/Mobile IP to authenticate itself by AAAH, to derive keys and to get the certificate of MAP. When the mobile node detects its fading RSS (Received Signal Strength) from current AP, it will trigger fast handoff before L2 handoff. The MN adds some authentication information into fast handover messages and sends them to PAR. Then, PAR adds its information and forwards these messages to NAR in pre-existing secure connection. NAR checks information in these messages and replies result to the mobile node through PAR in secure connection, finally PAR will forward this message to the mobile node. After these steps, the mobile node completes its transient authentication process. Next, we show the message flow and explain the details.

Figure 34 shows the total message flow in detail. We list extra information in RtSolPr, HI, HAck, PrRtAdv, and FNA for transient authentication as follows.

RtSolPr: $\{Ek_{MN-GK}\{CA_{MAP} \| CR \| HIP_{MN} \| T_{MN}\}, MAC_{MN}, R_{MN}\}Ek_{MN-PAR}$

$CR = H_1(GK, MAC_{MN}, NewCoA, T_{MN})$ GK: Group Key

RtSolPr contains credential CR generated by the roaming MN and a certificate of MAP identified by NAR. HIP_{MN} and MAC_{MN} are stored in authentication table of NAR. Other information is for challenge/response. These values are encrypted by keys.

HI: $\{Ek_{MN-GK}\{CA_{MAP} \| CR \| HIP_{MN} \| T_{MN}\}, MAC_{MN}, R_{PAR}\}$

PAR adds its challenge random number to identify the NAR for later use. HI contains information from RtSolPr and random number of PAR. This message is transmitted through existed security association.

HAck: $\{Ek_{MN-GK}\{T_{MN} \| CR' \| R_{NAR}\}, R_{PAR}\}$ **CR'**: Expire time for CR

When NAR receives the HI message, it checks the new care-of address of the mobile node and decides to register if it is allowed. NAR extracts the information of HI, then it checks the CAmap to authenticate the mobile node. Finally, NAR creates a new entry for the mobile node and records a temporary certificate called CR. At this time, the packets destined to MN through NAR will be forwarded because of its *temporary* authentication. NAR gives this entry a lifetime and passes the authentication of MN for a short period of time. This entry will be deleted if it expires. The lifetime is refreshed by receiving a correct CR sent by MN. Packets toward MN through NAR are dropped if this entry expires or it does not exist. So, HAck contains the lifetime of CR and a random number generated by NAR for later use. This message is transmitted to PAR in the secure association between PAR and NAR.

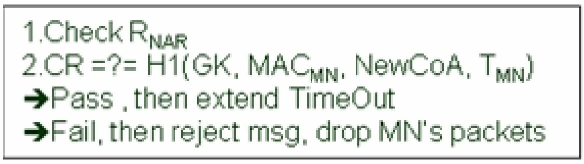
PrRtAdv: $\{Ek_{MN-GK}\{T_{MN} \| CR' \| R_{NAR}\}, R_{MN}\}Ek_{MN-PAR}$

PAR receives the HAck message and checks the R_{PAR} to identify NAR against replay attack. Then, it sends a PrRtAdv message to the MN. PrRtAdv contains the information generated by NAR. Two important values are CR' and R_{NAR}. This message is encrypted by keys. The MN stores these two values for later use.

FNA: $Ek_{MN-GK}\{R_{NAR} \| CR \| HIP_{MN}\}$

After layer2 handoff, the mobile node will send a FNA to NAR for receiving data packets. The MN sends R_{NAR} and CR to NAR. R_{NAR} is for challenge/response and CR is for extending the entry lifetime of mobile node in the authentication table of NAR. Then, MN performs re authentication after MIP/MAP registrations by sending the correct certificate CAmap. At this time, it needs 2-way handshaking to complete the re authentication because its correct certificate was derived from NAR previously. NAR operations upon receiving a FNA are shown in Fig. 35.

When NAR receives the FNA message, it extracts the R_{NAR} and CR to check whether they are correct or not. If values are correct, it extends the entry lifetime. If the authentication is failed, packets toward the MN are dropped. The mobile node can derive a buffer time to receive data packets quickly through transient authentication. Figure 36 shows the message flow of transient authentication.



1.Check R_{NAR}
2.CR =?= H1(GK, MAC_{MN}, NewCoA, T_{MN})
➔Pass , then extend TimeOut
➔Fail, then reject msg, drop MN's packets

Fig. 35: NAR Operations.

When NAR receives the FNA message, it extracts the R_{NAR} and CR to check whether they are correct or not. If values are correct, it extends the entry lifetime. If the authentication is failed, packets toward the MN are dropped. The mobile node can derive a buffer time to receive data packets quickly through transient authentication. Figure 36 shows the message flow of transient authentication.

3.4 Comparison

There is another authentication scheme called L3-FHR [10]. The main idea is to broadcast authentication reply packets to all L3-FHR APs. So, MN can re-authenticate with new target AP without authenticating with home AAA server again. It may reduce the re-authentication time. New target AP will stop dropping packets after MN completes the re-authentication. Its message flow is shown on left-side in Fig. 37. We put our proposed scheme on right-side in Fig. 37 to show the difference.

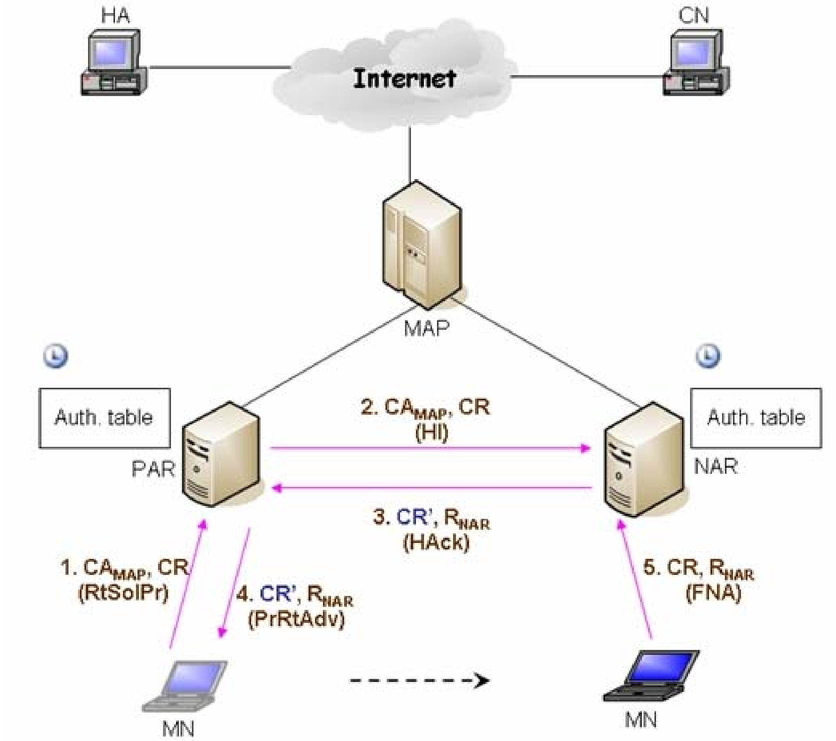


Fig: 36: Scenario of Transient Authentication Message Flow.

Then, we use a table to list the difference between these two schemes in detail, as shown in Fig. 38. Both these two schemes are to pre-send the authentication information to the new target AP before handoff. But the two-stage scheme not only pre-sends this authentication information but also registers its authentication information temporarily with new target AP before L2 handoff. It can stop dropping packets on new target AP temporarily. In other words, L3-FHR AP will drop packets if MN doesn't complete the re-authentication process. L3-FHR duplicates authentication packets to all L3-FHR APs, but our two-stage scheme sends authentication packets through fast handover signaling. The following are the main differences.

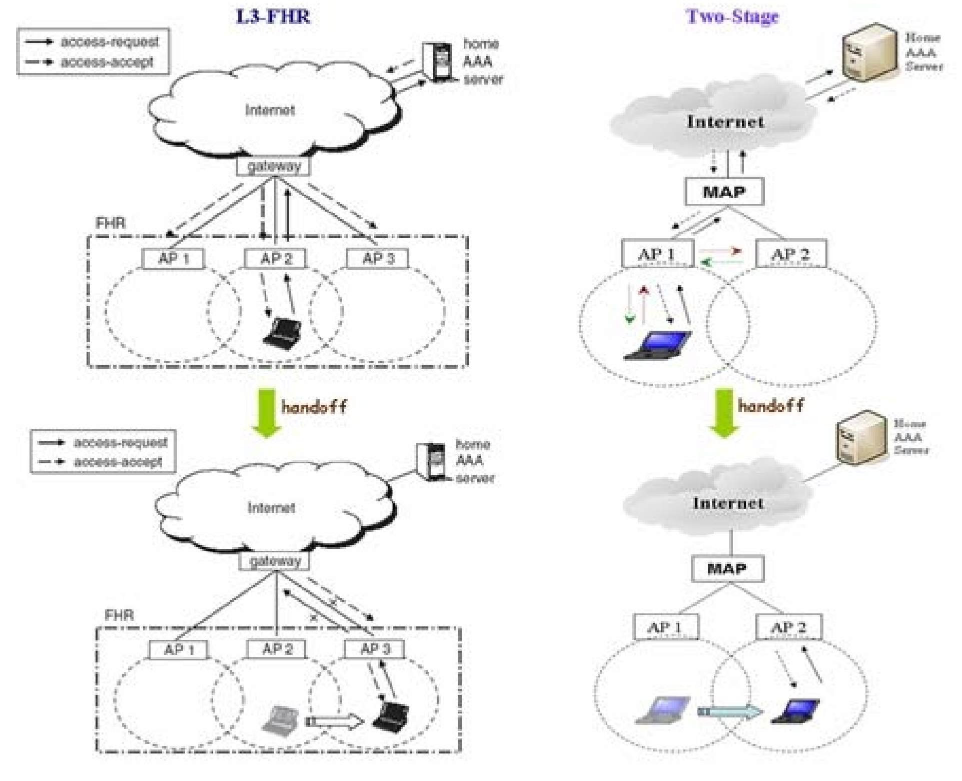
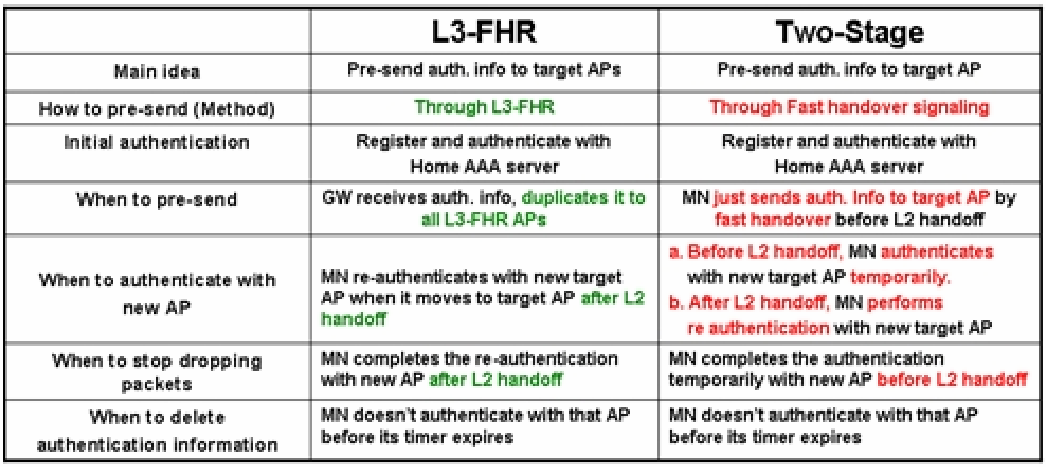


Fig. 37: L3-FHR & Two-Stage Message Flow.



	L3-FHR	Two-Stage
Main idea	Pre-send auth. info to target APs	Pre-send auth. info to target AP
How to pre-send (Method)	Through L3-FHR	Through Fast handover signaling
Initial authentication	Register and authenticate with Home AAA server	Register and authenticate with Home AAA server
When to pre-send	GW receives auth. info, duplicates it to all L3-FHR APs	MN just sends auth. Info to target AP by fast handover before L2 handoff
When to authenticate with new AP	MN re-authenticates with new target AP when it moves to target AP after L2 handoff	a. Before L2 handoff, MN authenticates with new target AP temporarily. b. After L2 handoff, MN performs re authentication with new target AP
When to stop dropping packets	MN completes the re-authentication with new AP after L2 handoff	MN completes the authentication temporarily with new AP before L2 handoff
When to delete authentication information	MN doesn't authenticate with that AP before its timer expires	MN doesn't authenticate with that AP before its timer expires

Fig. 38: L3-FHR & Two-Stage Comparisons.

3.5 Contributions

We list the contribution of two-stage authentication scheme as follows.

- Two-stage authentication
  - Early to derive the access right in the new domain
  - Perform transient authentication before L2 handoff
- Fast handover signaling piggybacks these authentication information
  - No protocol overhead
  - Simple and feasible implementation
- Reduce original IEEE 802.11 authentication procedure time
  - Early to send “identity” to new BS (AP)
- Main factors affecting packet loss rate
  - Authentication processing time (100ms v.s 300ms)
    - ✓ UDP Packet loss rate

        - ➢ 100ms & 1Mbps
          41.6% (no-transient) ➔ 13.6% (transient)
        - ➢ 300ms & 1Mbps
          100% (no-transient) ➔ 34.4% (transient)
            - UDP data sending rate (100Kbps v.s 1Mbps)
    - ✓ UDP Packet loss rate
        - ➢ 32% (100kbps & no-transient) ➔ 8% (100kbps & transient)
        - ➢ 41.6% (1Mbps & no-transient) ➔ 13.6% (1Mbps & transient)
- ● More efficient than L3-FHR authentication scheme
    - Just send authentication information to only one target AP instead of all L3-FHR APs.

3.6 Comments

Many authentication methods are proposed in different goals. Some focus on layer 2 authentication and the others focus on upper layers. Each layer may have its authentication function. A mobile node needs to perform many authentication procedures in each layer if we want to have a secure communication. Generally, the MN needs to perform layer 2 and user authentication when it roams. Currently, IEEE 802.1x is a layer 2 authentication method, it takes approximately 1200ms or longer to complete total steps. If a mobile node supports this function with Mobile IP, it may cause some problems. Figure 39 shows the message flow with IEEE 802.11, IEEE 802.11i and Mobile IP.

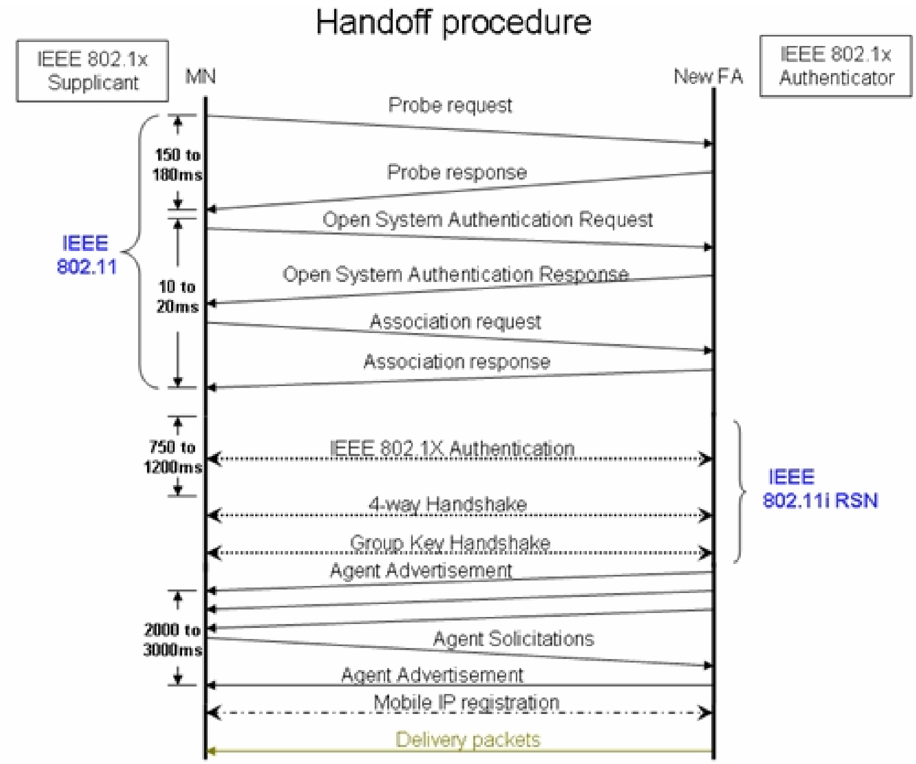


Fig. 39: Handoff Procedure with IEEE 802.11 & IEEE 802.11i & MIP.

A mobile node in foreign domain needs to register with home agent periodically before its lifetime expires. If layer 2 authentication processing time is longer than registration lifetime, the home agent begins to drop packets destined to MN. But a mobile node needs to complete layer 2 and layer 3 handoffs when it roams, this takes much longer time to finish these operations excluding authentication processes. Mobile IP registration lifetime would be increased if we add authentication functions. This is a tradeoff between fast handoff and secure authentication. So, transient authentication is an essential component if authentication is used during handoff.

3.7 Performance Evaluation

In order to verify the performance of the transient authentication mechanism proposed in Chapter 3, we use ns-2 distribution version *ns-allinone-2.1b7a* patched with two available modules, NO Ad-Hoc Routing Agent (NOAH) [12] and HMIPv6 with Fast-handover (FHMIP) [13]. We add some extra features and modify some codes based on these modules, including re (user) authentication mechanism and transient authentication with fast handover protocol described in Chapter 3.

We build up the network topology to evaluate the proposed mechanism as shown in Fig. 40. There are nine nodes in the network topology, including Correspondent Node (CN), Home Agent (HA), Mobility Anchor Point (MAP), Previous Access Router (PAR), New Access Router (NAR), Mobile Node (MN), and three fixed nodes. The bandwidth (Megabits/second) and link delay (milliseconds) are shown on the link between two nodes. The distance between two access routers is 70 meters and each router advertisement interval is one per second [14]. The wireless coverage area of the access router is approximately 100 meters in radius. We set up the handoff delay to 20ms in our simulation. The mobile node starts to move from PAR to NAR with speed 1 meter per second at the 10th second. The total simulation period is 80 seconds.

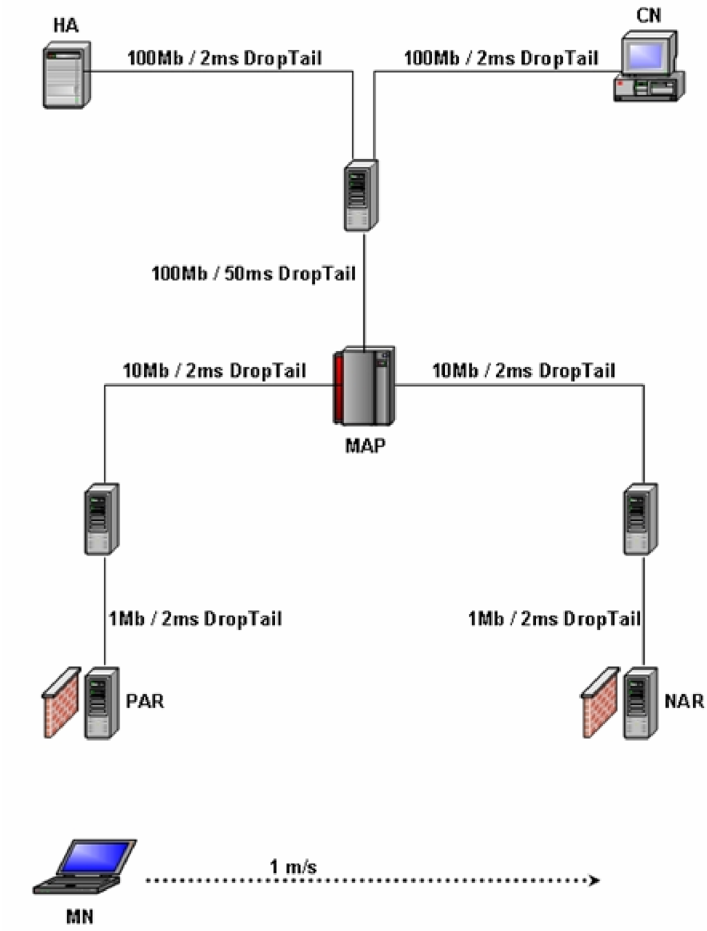


Fig. 40: Network Topology for Simulation.

We evaluate three experiments in this topology. First, A TCP sender (CN) starts to send packets at 5s until the end of the simulation and we would like to observe its variance when using FTP service. Second, A UDP sender (CN) starts to send packets at 5s until the end of the simulation. Also, we could observe the different impacts between UDP and TCP packets. Finally, we want to know how the authentication time

affects the performance when it increases. We set the *L2 handoff time* to 20ms. The default time value of sending authentication information by either the sender or receiver is 100ms. We adjust this value to observe the variance in different processing time.

3.8 Experiment and results

TCP Experiment

First, we see the different handoff time between original Mobile IP and HMIPv6 with Fast handover, as shown in Fig. 41. The original flat Mobile IP takes approximately 4 seconds to perform the handoff from 50 to 54 sec. This time is too long so it interrupts the TCP connection between CN and MN, hence it can't provide good services such as Multimedia streaming or VOIP. HMIPv6 with Fast handover encounters handoff at 40.6 sec and it takes 100ms to continue the reception of the packets from new access router at 40.7sec. So, this handoff time is better for providing Multimedia streaming or VOIP service. Based on this advantage of Fast handover, we try to add an extra authentication mechanism on it. If authentication mechanism is added to the fast handover mechanism, we should control the overhead time in 50ms. Next, we show three cases under fast handover, including original fast handover, fast handover without transient authentication, and fast handover with transient authentication. These three cases feature differences on handoff as shown in Fig. 42.

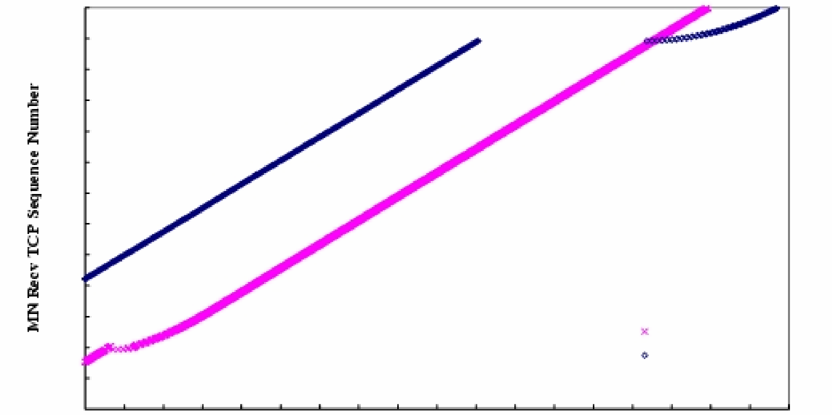


Fig. 41: FHMIP and Flat Mobile IP without Authentication.

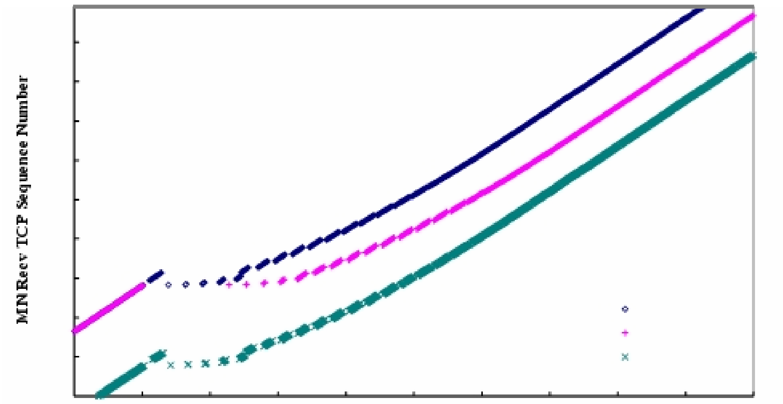


Fig. 42: Original & with & without Transient Auth. under Fast Handover.

We know that transient authentication reduces the handoff time, this is same as original structure under fast handover as shown in Fig. 42. Also, the data curve for fast handover with transient authentication is almost the same as fast handover without extra authentication mechanism, this result matches our expectation.

The mobile node can't receive any data packets from 40.51 second to 41.14 second during fast handoff and re (user) authentication as shown in Fig. 43. It takes a mobile node 630ms penalty for continuing to receive the data packets destined for itself. The time penalty may break the TCP connection if the delay increases. The delay time causes the mobile node to defer the reception of data packets. The new Access Router drops 22 data packets destined for the MN during authentication period, and 8 data packets are lost during fast handoff period. So, the total lost packets toward the mobile node are 30 data packets. In this experiment, we set the *L2 handoff time* value to 20ms. If we add more delay time, more packets will be dropped, these data packets dropped due to handoff and authentication will be retransmitted by the correspondent node.

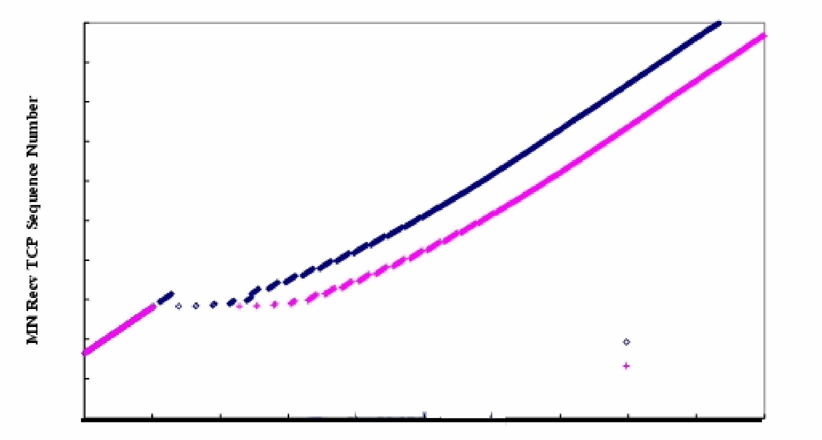


Fig. 43: Authentication Processing Time 100ms.

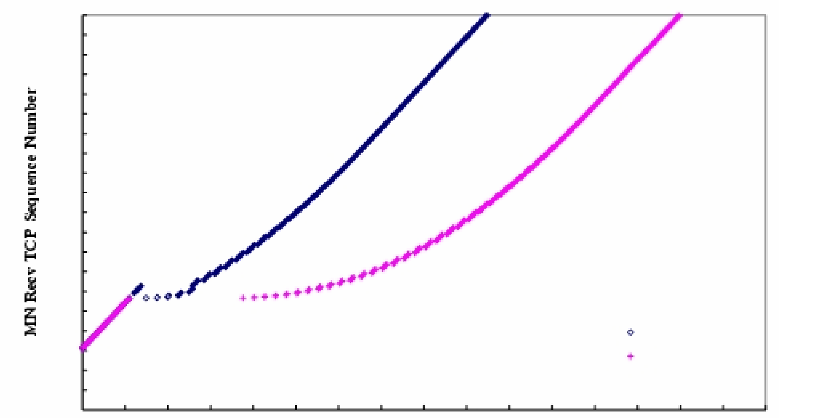


Fig. 44: Authentication Processing Time 200ms.

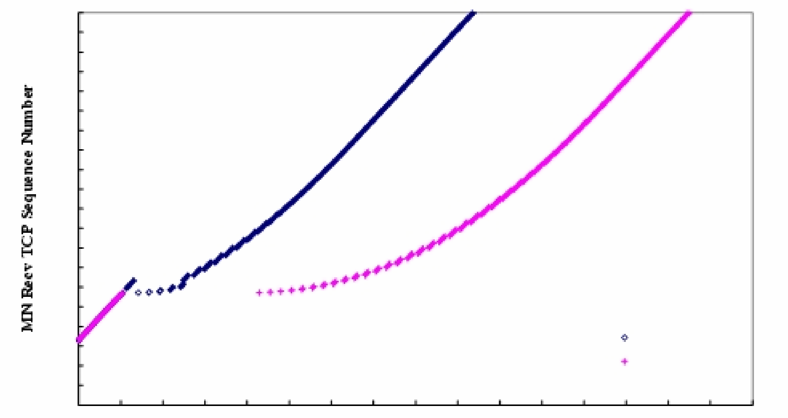


Fig. 45: Authentication Processing Time 300ms.

Performing transient authentication mechanism could decrease the packet loss rate during the authentication process as shown in Fig. 43. Transient authentication mechanism can alleviate the delay time as well, this allows a mobile node to receive data packets earlier. Transient authentication mechanism offers a buffering time for a mobile node to pass temporary authentication and receive data packets quickly. Also, the mobile node uses temporary certificate to extend the authentication time not to expire when the mobile node doesn’t complete the re (user) authentication. So, no data packets will be dropped during authentication process. The mobile node gets a temporary access right to keep receiving data packets in the new domain such as NAR. Compared with the scheme without transient authentication as shown in Fig. 43, formal authentication indeed increases or enlarges the delay time and packet loss rate. If we perform transient authentication prior to handoff, the performance will be improved. Next, we show the growth of lost packets if we increase the authentication process time.

The delay of receiving packets is indeed higher depends on processing time as shown in Figu. 43, 44 and 45. We combine three different cases including with and without transient authentication into two graphs shown below. The authentication processing time is 100ms, 200ms and 300ms, so the total completion time is 330ms, 700ms and 1sec respectively, as shown in Fig. 47 and 48.

Authentication processing time	Number of lost packets
100 ms	30
200 ms	31
300 ms	32

Table 3: Authentication Processing Time & Lost Packets Relationship.

We can observe that the number of lost TCP packets increases slowly if we increase authentication processing time quickly. This result is due to TCP sliding window effect. The mobile node can't receive the packets dropped by Access Point if it doesn't authenticate with that Access Point. Therefore, the sender can't receive the ACKs from receiver. The packets of offered window may be dropped by Access Point due to authentication mechanism. Figure 46 shows the TCP sliding window.

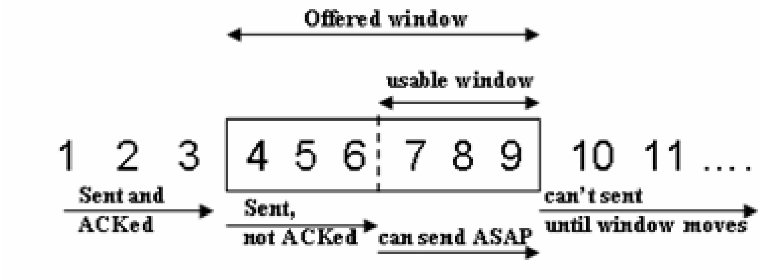


Fig. 46: TCP Sliding Window.

Sender will retransmit the first packet of offered window if the packets of usable window are sent and the sender doesn't receive any ACKs from receiver. The authentication processing time increases the delay for the mobile node to receive the packets and it may interrupt the TCP connections between the CN and the MN.

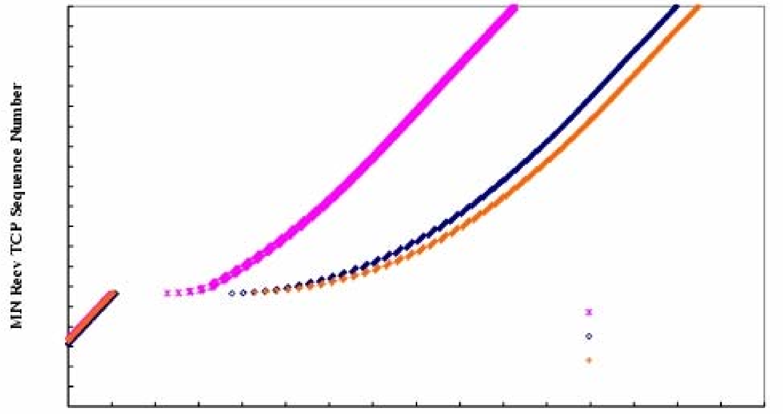


Fig. 47: 100ms - 300ms Cases without Transient Authentication.

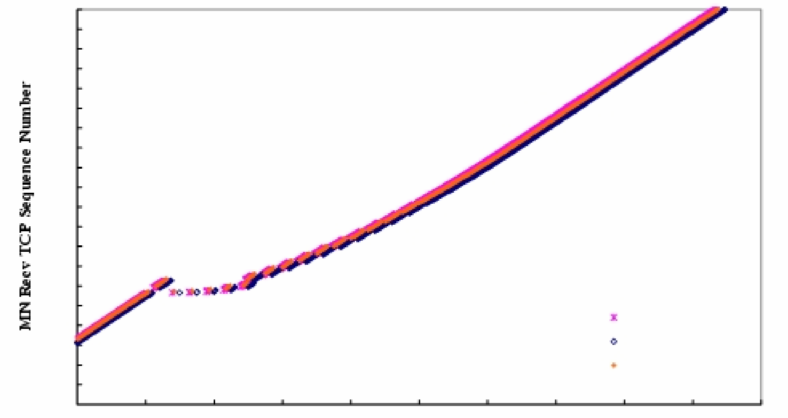


Fig. 48: 100ms - 300ms Cases with Transient Authentication.

UDP Experiment

We can observe the performance on TCP as mentioned above. Now we change the service to UDP, the result is presented as follows. We set the UDP related traffic value in the following:

$cbr set type_ CBR
$cbr set packet_size_ 1000
$cbr set rate_ 1mb
$cbr set random_ false

We will discuss and compare the effect of various data rates on UDP services. UDP service is different from TCP service because it is a connectionless service, so the lost UDP packets will not be retransmitted again. If the packet loss rate is high, there will be a noticeable gap, as shown in the following figures.

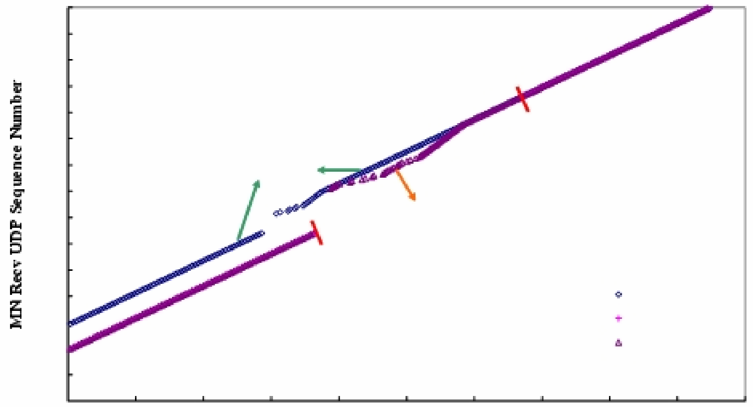


Fig. 49: Original & with & without Transient Auth. under Fast Handover.

Figure 49 shows the difference between three cases under UDP services. We could observe that the transient authentication curve is similar to the original fast handover. So, the packet loss rate is almost equal, while the packet loss rate of fast handover without transient authentication is high.

no-transient timestamp	no-transient udp seqno
40.429763	442
41.069763	450

Table 4: No-transient with UDP 100Kbps during handoff.

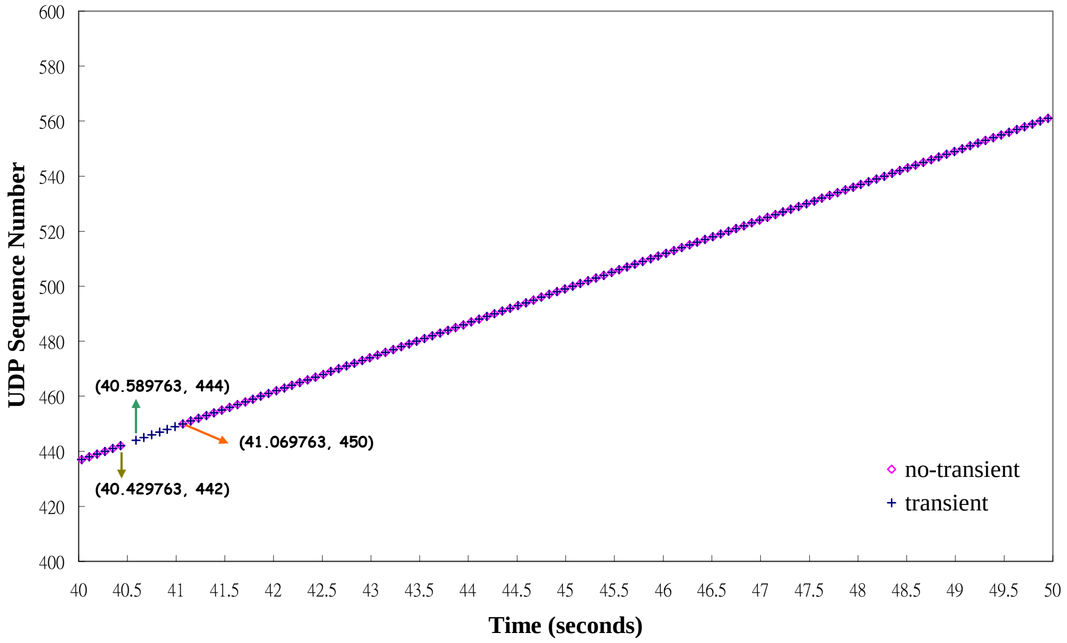


Fig. 50: UDP with 100Kbps Data Rate.

In order to analyze the UDP packet loss rate, we add a sequence number to each packet to observe the growth of packet loss. First, we set the UDP data rate to 100kbps as shown in Fig. 50, and there are 8 lost packets during re-authentication without transient authentication in 640ms. This elapsed time is too large to offer a good Multimedia Streaming or VOIP service. Actually, when users of a multimedia streaming application move from the coverage area of an AP (access point) to the other, the connection must be handed off in approximately 150 milliseconds, otherwise the user will feel the jitter affect. So, we use the transient authentication during handoff to reduce the authentication time in 150ms, as shown in Fig. 50. It can improve the QoS for multimedia streaming application.

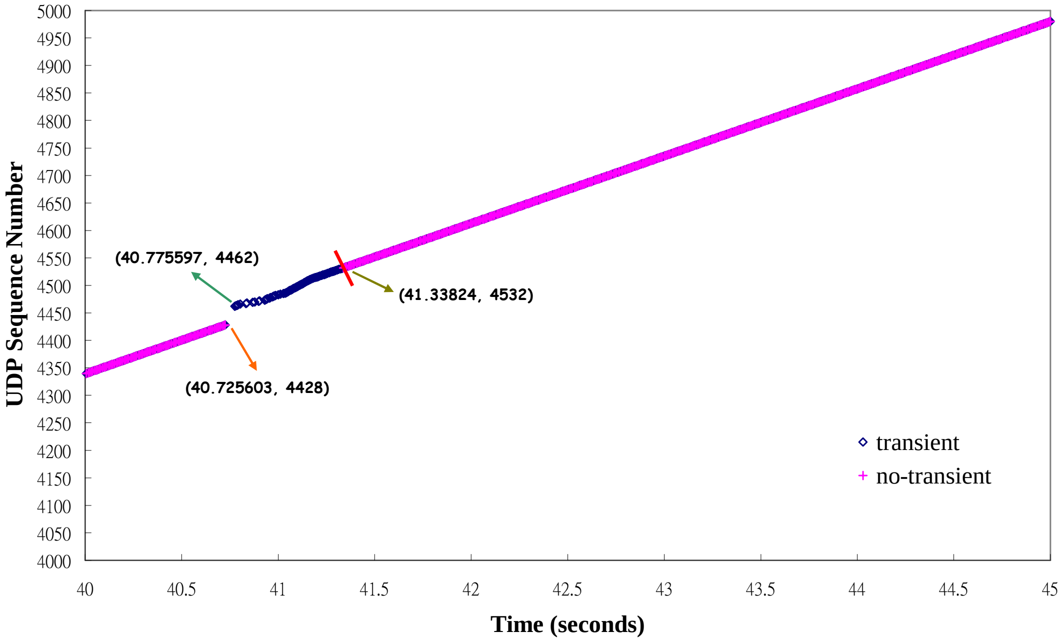


Fig. 51: UDP Data Rate 1Mbps.

no-transient timestamp	no-transient udp seqno
40.725603	4428
41.33824	4532

Table 5: No-transient with UDP 1Mbps during handoff.

We increase the data rate to 1Mbps to observe the difference. First, we address the fast handover without transient authentication case. At 40.73sec, the mobile node receives UDP sequence number 4428 and continues to receive UDP packets at 41.34sec with UDP sequence number 4532. So, the total number of lost packets is 104 in 0.6sec, we can see that there is a very large gap shown in Fig. 51. Compared with TCP experiment, the packet loss rate of UDP experiment is higher because TCP will adjust its window to slow down its sending rate. UDP keeps its constant sending rate at 1Mb, so the packet loss rate is higher if the mobile node doesn't perform transient authentication mechanism to get a temporary certificate to pass the authentication in the new domain. We try to use the fast handover protocol to reduce the handoff time, however the authentication process still causes a significant delay which is approximately 330ms or even higher. We can realize that the authentication process does affect the handoff performance drastically if we use more complex authentication mechanism. Then, we calculate the packet loss rate in 2 seconds during handoff period as follows.

Sending rate	Packet drops		Packet loss rate	
100kbps	8	2	**32%(no-transient)**	**8%(transient)**
1Mbps	104	34	**41.6%(no-transient)**	**13.6%(transient)**

Table 6: UDP Packet Loss Rate (no-transient & transient).

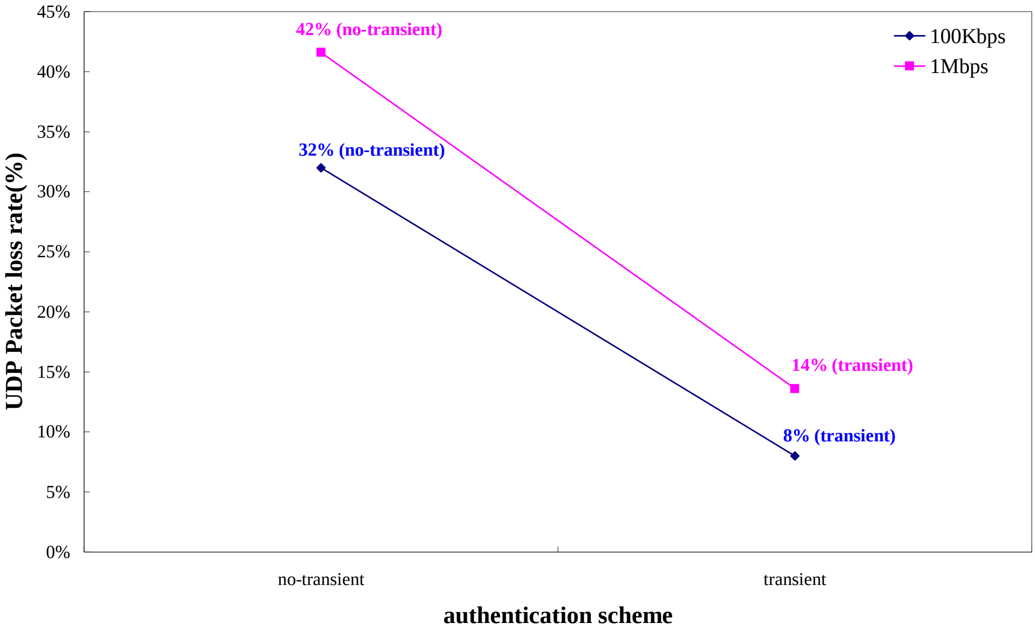


Fig. 52: Packet Loss Rate with Different UDP Data Rate.

Figure 52 show that the UDP data rate and transient authentication indeed affect the packet loss rate. Next, we show the final figure regarding increasing the authentication time in 300ms under UDP service in data rate 1Mbps.

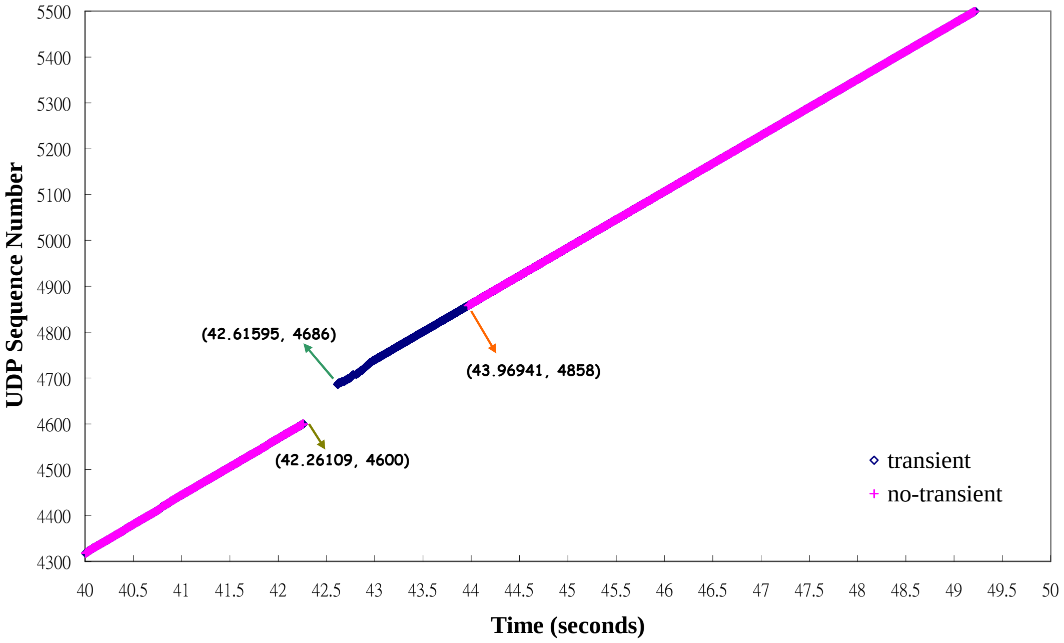


Fig.53: Authentication Processing time.

Figure 53 shows that the authentication processing time also increase the packet loss rate.

Auth. processing time	Packet drops	Packet loss rate
100ms (no-transient)	104	**41.6%**
100ms (transient auth)	34	**13.6%**
300ms (no-transient)	258	**100%**
300ms (transient auth)	86	**34.4%**

Table 7: Packet drops with authentication processing time 100ms & 300ms, sending rate 1Mbps.

3.9 Section Conclusion

In this project, we investigated a two-stage authentication scheme which includes transient authentication and re-authentication mechanism. The user authentication is called *re-authentication*. The re-authentication signaling consists of 4-way handshaking and we use it to simulate the authentication time during handoff period. In the original structure of Mobile IP, it needs approximately 3-4 sec to complete the handoff process excluding authentication. If we add authentication process on it, the handoff time will increase drastically due to the complex authentication mechanism. Since a more complex authentication mechanism needs longer time to process, the transient authentication becomes important if we try to reduce the authentication time during handoff period. In our experiment as discussed in section 3.8, we demonstrate that the packet loss rate increases when UDP sending rate increases. Packet loss rate is reduced to 8% with transient authentication when UDP sending rate is 100Kbps. Packet loss rate is reduced to 13.6% when UDP sending rate is 1Mbps. If we combine transient authentication with fast handover protocol, it is able to reduce the packet loss rate and perform well as original fast handover protocol without authentication mechanism. It has no protocol overhead and is feasible to implement. The re-authentication signaling is mainly to simulate an authentication process, and we can understand how the packet loss rate will change if we add the authentication mechanism in this project. Finally, the proposed transient authentication method piggybacks on authentication information through fast handover protocol without additional signaling overhead. We know that IEEE 802.1x is a MAC layer authentication mechanism. It takes more than 1200ms to complete the authentication process. If we count the total disconnection time such as scan, authentication and association, the time value is too large to be acceptable for certain applications. Nowadays, many proposed methods are trying to improve the performance in different phases. Even if we use the best available method in each phase, the total disconnection time is still too large to offer a high Quality of Service for VOIP or Multimedia streaming. So, the tradeoff between authentication and fast handoff is difficult. In this project, we perform the user authentication to open the filtering table for MN on new access router, which will drop or forward the packets toward MN according to the authentication table. We use transient authentication concept early to get the access right in the new domain when MN roams to it based on fast handover protocol.

In order to provide the functions of authentication and security, definitely, the original handoff time will be increased, this in turn increases packet loss rate and degrades the quality of service. So, how to maintain good service under the framework of providing authentication and security will be an essential issue in the future.

The proposed method in this project is designed under Mobile IP architecture. It is a user authentication mechanism in which we send the identity of the mobile node and user authentication information before handoff really starts, this could reduce the total authentication time during the handoff process. Also, we may modify the layer 2 authentication of IEEE 802.1x to support the pre re-authentication function or enhance its security and key distribution, this will be studied in the future research.

五、結論與討論

The speedy handover mechanism is proposed to meet one of the objectives of this subproject. It can enhance the performance of wireless handover. Different from the famous fast handover scheme that a mobile node must discover the movement by itself, an access point take charge of detecting the movements of mobile nodes in speedy handover mechanism. As a result, the proposed scheme can work no matter the radio covering area of the two neighbored FAs are overlapped or not. We are planning to refine the speedy handover further and submit it to an international conference.

In the second year of this subproject, we will develop a scheme for location management of mobile hosts. Load balance and QoS provision in wireless access networks will be the targets of our new scheme.

In this project, the concept of “discrete scan” is newly proposed. Based on the discrete scan scheme, the next AP in a handoff can be discovered prior to the occurrence of handoffs. With the survey of fast handoff schemes based on the prediction of next APs, a series of handoff schemes, including context caching, proactive key distribution, buffering and forwarding, inter layers handoff design, are suitable to cooperate with discrete scan scheme to fasten layer-2 and layer-3 handoff to achieve the goal of seamless handoffs.

The principle of discrete scan lies on the decomposition of the passive scan procedure in a traditional layer-2 handoff into discrete pieces of sniffing periods prior to the occurrence of handoff. A mobile node utilizes the possible idle time of its NIC as the sniffing periods so that the disruption caused by a sniffing activity is controlled less than 50 ms to fulfill the minimal requirement of seamless handoff features.

With the application of discrete scan scheme, a mobile node can further select an appropriate next AP among the next-AP candidates with help of information extracted from MAC headers in those frames collected in sniffing periods. Several mechanisms to select the next AP with desired features are proposed in this project, such as the ratio of stations locating within coverage of the mobile node to that in the BSS and the station count discovered in the last sniffing period before a handoff are initiated to select the nearest AP as the next AP. Besides, a scheme to identify an approached AP and an indicator to predict the available bandwidth of concerned BSS are briefly discussed in this project.

With the benefits of simplicity and nearly real-time characteristic, the station count mechanism is elaborated and evaluated. The setting of a sniffing period for discrete scan scheme, performance for the mechanism to hit the nearest AP as well as the impact on service QoS caused by the absence from working channel in sniffing periods for discrete scan scheme in a DCF wireless environment are demonstrated with both analytical model and numerical results. For further verification, those for EDCA wireless environment are analyzed with results generated in ns-2 simulations.

The numerical as well as the simulation results show that discrete scan with station count mechanism provides considerable high hit ratio on the selection of the nearest AP among the next-AP candidates discovered by a mobile node. Although the imbalance of number of stations in the next-AP candidates is still an inevitable factor to affect the selection accuracy for the nearest AP, its effects can be significantly reduced by setting a proper length of sniffing periods.

The nature that an AP, responsible for all downlink traffic, has the same transmission opportunity as an arbitrary station, taking care only its own uplink traffic, in a BSS may causes insufficient downlink bandwidth for a symmetric bidirectional connection such as a VoIP application. The downlink problem for VoIP traffic is supposed to resolve with the piggyback scheme reported in 802.11e standard, because it is an inherent QoS problem instead of being caused by discrete scan scheme. However, the impact on the service QoS caused by discrete scan scheme is discussed in view of uplink flow for VoIP traffic. Simulation results show the induced disruptions being maintained within a tolerable level in the EDCA environment for all cases of reasonable number of stations in a BSS; However, numerical analysis shows a mobile node with discrete scan scheme in DCF wireless LAN can sustain acceptable QoS when less than six stations are in a BSS.

In the future, we will further improve the scheme by investigating a scenario that a mobile VoIP node with discrete scan scheme works in EDCA environment. The observation will focus on the QoS degradation induced by discrete scan scheme. Besides, an integrated selection mechanism with sniffing function and an algorithm to identify a station via reception of its transmitting frame will be developed. Finally, we will plan to implement a complete set of discrete scan as well as selection mechanism into a real WiFi phone. The WiFi phone shall be proved with experiments the capability of fast layer-2 handoff because latency of probe phase has been eliminated. The performance to select a desired next AP shall be verified with experiments. Eventually, the strategies addressed for fast handoff in the related works, such as context caching, buffering-and-forwarding and inter layer handoff design will be implemented into the APs and FAs to cooperate with a mobile node with discrete scan scheme, so that a complete wireless environment can be configured to support seamless inter-domain handoff for a next generation WiFi services.

In order to provide the functions of authentication and security, definitely, the original handoff time will be increased, this in turn increases packet loss rate and degrades the quality of service. So, how to maintain good service under the framework of providing authentication and security will be an essential issue in the future.

The proposed method in this project is designed under Mobile IP architecture. It is a user authentication mechanism in which we send the identity of the mobile node and user authentication information before handoff really starts, this could reduce the total authentication time during the handoff process. Also, we may modify the layer 2 authentication of IEEE 802.1x to support the pre re-authentication function or enhance its security and key distribution, this will be studied in the future research.

六、參考文獻

[1] IEEE 802.11b, IEEE Std. 802.11-1999.

[2] The Network Simulator – ns2, http://www.isi.edu/nsnam/ns.

[3] M. Shin, A. Mishra, and W. Arbaugh, “Improving the Latency of 802.11 Hand-offs Using Neighbor Graphs,” ACM Mobisys, 2004.

[4] C. C. Tseng, L. H. Yen and H.H. Chang “Topology-Aided Cross-Layer Fast Handoff Designs for IEEE 802.11 Mobile IP environment,” IEEE Communications Magazine, Dec. 2005

[5] G. Bianchi, “Performance Analysis of the IEEE 802.11 Distributed Coordination Function,” IEEE Journal on selected areas in communications, Vol.18 No.3, March 2000

[6] Ghuangxiong Guo, Zihua Guo, Qian Zhang and Wenwu Zhu, “A seamless and Proactive End-to-End Mobility Solution for Roaming Across Heterogeneous Wireless Networks” IEEE journal on selected area in communication, vol. 22, No, June 2004.

[7] IEEE Std. 802.11e, IEEE computer society, November 2005.

[8] http://www.isi.edu/nsnam/ns/

[9] http://www.tkn.tu-berlin.de/research/802.11e_ns2/

[10] S. Pack and Y. Choi, “Fast handoff scheme based on mobility prediction in public wireless LAN systems,” IEE Proceedings, Vol. 151, No.5, October 2004

[11] Jyh-cheng chen et al., “Wireless LAN Security and IEEE 802.11i,” IEEE Wireless
Communications, February 2005

[12] Jörg Widmer, “Extensions to the ns Network Simulator (NOAH),” http://www.informatik.uni-mannheim.de/pi4/projects/MobileIP/ns-extension/

[13] Robert Hsieh, “fhmip ns2-extension,” http://mobqos.ee.unsw.edu.au/~robert/nsinstall.php#beginning

[14] C. Perkins, “IP Mobility Support,” RFC 2002, IETF, October 1996Nicolas Montavont and Thomas

[15] Noël LSIIT, “Handover Management for Mobile Nodes in IPv6 Networks,” IEEE Communications Magazine, August 2002

[16] Robert Hsieh, Aruna Seneviratne, Hesham Soliman, Karim El-Malki, “Performance analysis on Hierarchical Mobile IPv6 with Fast-handoff over End-to-End TCP,” Proceedings of GLOBECOM, Taipei, Taiwan 2002.

[17] Arunesh Mishra, Minho Shin, Willian Arbaugh, “An Empirical Analysis of the IEEE 802.11 MAC Layer handoff process”, University of Maryland Technical Report, UMIACS-TR-2002-75, 2002

[18] Sangho Shin, Anshuman Singh Rawat, Henning Schulzrinne, “Reducing MAC Layer Handoff Latency in IEEE 802.11 Wireless LANs,” MobiWac'04, October 1, 2004

[19] Fang Feng, Douglas S. Reeves, “Explicit Proactive Handoff with Motion Prediction for Mobile IP,” WCNC'04, 2004

[20] M.S. Bargh, R.J. Hulsebosch, E.H. Eertink, A. Prasad, H. Wang, P. Schoo, “Fast Authentication Methods for Handovers between IEEE 802.11 Wireless LANs,” WMASH’04, October 1, 2004

[21] IEEE Trial-Use Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation, IEEE Std 802.11f, Jul. 2003.

[22] Wei-Min Yao, Yaw-Chung Chen, “An Enhanced Buffer Management Scheme for Fast Handover Protocol.” Proceedings of 24th International Conference Distributed Computing Systems Workshops, Pages: 896 – 901, 23-24 March 2004

[23] S. Seshan et al., “Handoffs in Cellular Wireless Networks: The Daedalus Implementation and Experience,” Kluwer J. Wireless Personal Communication, vol. 4, No. 2, pp. 141-162, March 1997

[24] E. Shim et al., “Low Latency Handoff for Wireless IP QoS with Neighborcasting,” in Proc. ICC

2002, April 2002.

[25] K. Malki et al., “Low Latency Handoffs in Mobile IPv4,” *Internet Draft*, IETF, draft-ietf-mobileip-lowlatency-handoffs-v4-04.txt, June 2002.

[26] C. Blondia et al., “Performance Evaluation of Layer 3 Low Latency Handoff Mechanisms,” Mobile Networks and Applications, pp. 633-645, 2004

[27] Robert Hsieh, Zhe Guang Zhou, Aruna Seneviratne, “S-MIP: A Seamless Handoff Architecture for Mobile IP,” IEEE INFOCOM 2003

[28] Ali Diab, Andreas Mitschele-Thiel, Esam Al Nasouri, René Böringer, Jingan Xu, “Mobile IP Fast Authentication Protocol,” Ilmenau University of Technology

[29] Ali Diab, Andreas Mitschele-Thiel, Jingan Xu, “Performance Analysis of the Mobile IP Fast Authentication Protocol,” MSWiM’04, October 4–6, 2004

[30] C. Rigney et al., “Remote Authentication Dial In User Service,” RFC 2865, IETF, June 2000

[31] Zhang Hong, He Rui, Yuan Man, Kan Zhigang, “A Novel Fast Authentication Method for Mobile Network Access,” International Conference for Young Computer Scientists (ICYCS), August 2003

[32] Pat R. Calhoun et al., “Diameter Mobile IPv4 Application,” Internet Draft, IETF, August 2004, draft-ietf-aaa-diameter-mobileip-20.txt